

Derivatives and Conditional Probability

SECTION 31. DERIVATIVES ON THE LINE*

This section on Lebesgue's theory of derivatives for real functions of a real variable serves to introduce the general theory of Radon–Nikodym derivatives, which underlies the modern theory of conditional probability. The results here are interesting in themselves and will be referred to later for purposes of illustration and comparison, but they will not be required in subsequent proofs.

The Fundamental Theorem of Calculus

To what extent are the operations of integration and differentiation inverse to one another? A function F is by definition an *indefinite integral* of another function f on $[a, b]$ if

$$(31.1) \quad F(x) - F(a) = \int_a^x f(t) dt$$

for $a \leq x \leq b$; F is by definition a *primitive* of f if it has derivative f :

$$(31.2) \quad F'(x) = f(x)$$

for $a \leq x \leq b$. According to the *fundamental theorem of calculus* (see (17.5)), these concepts coincide in the case of continuous f :

Theorem 31.1. *Suppose that f is continuous on $[a, b]$.*

(i) *An indefinite integral of f is a primitive of f : if (31.1) holds for all x in $[a, b]$, then so does (31.2).*

(ii) *A primitive of f is an indefinite integral of f : if (31.2) holds for all x in $[a, b]$, then so does (31.1).*

*This section may be omitted.

A basic problem is to investigate the extent to which this theorem holds if f is not assumed continuous. First consider part (i). Suppose f is integrable, so that the right side of (31.1) makes sense. If f is 0 for $x < m$ and 1 for $x \geq m$ ($a < m < b$), then an F satisfying (31.1) has no derivative at m . It is thus too much to ask that (31.2) hold for all x . On the other hand, according to a famous theorem of Lebesgue, if (31.1) holds for all x , then (31.2) holds almost everywhere—that is, except for x in a set of Lebesgue measure 0. In this section *almost everywhere* will refer to Lebesgue measure only. This result, the most one could hope for, will be proved below (Theorem 31.3).

Now consider part (ii) of Theorem 31.1. Suppose that (31.2) holds almost everywhere, as in Lebesgue's theorem, just stated. Does (31.1) follow? The answer is no: If f is identically 0, and if $F(x)$ is 0 for $x < m$ and 1 for $x \geq m$ ($a < m < b$), then (31.2) holds almost everywhere, but (31.1) fails for $x \geq m$. The question was wrongly posed, and the trouble is not far to seek: If f is integrable and (31.1) holds, then

$$(31.3) \quad F(x+h) - F(x) = \int_a^b I_{(x, x+h)}(t) f(t) dt \rightarrow 0$$

as $h \downarrow 0$ by the dominated convergence theorem. Together with a similar argument for $h \uparrow 0$ this shows that F must be continuous. Hence the question becomes this: If F is continuous and f is integrable, and if (31.2) holds almost everywhere, does (31.1) follow? The answer, strangely enough, is still no: In Example 31.1 there is constructed a continuous, strictly increasing F for which $F'(x) = 0$ except on a set of Lebesgue measure 0, and (31.1) is of course impossible if f vanishes almost everywhere and F is strictly increasing. This leads to the problem of characterizing those F for which (31.1) does follow if (31.2) holds outside a set of Lebesgue measure 0 and f is integrable. In other words, which functions are the integrals of their (almost everywhere) derivatives? Theorem 31.8 gives the characterization.

It is possible to extend part (ii) of Theorem 31.1 in a different direction. Suppose that (31.2) holds for every x , not just almost everywhere. In Example 17.4 there was given a function F , everywhere differentiable, whose derivative f is not integrable, and in this case the right side of (31.1) has no meaning. If, however, (31.2) holds for every x , and if f is integrable, then (31.1) does hold for all x . For most purposes of probability theory, it is natural to impose conditions only almost everywhere, and so this theorem will not be proved here.[†]

The program then is first to show that (31.1) for integrable f implies that (31.2) holds almost everywhere, and second to characterize those F for which the reverse implication is valid. For the most part, f will be nonnegative and F will be nondecreasing. This is the case of greatest interest for probability theory; F can be regarded as a distribution function and f as a density.

[†]For a proof, see RUDIN₂, p.179.

In Chapters 4 and 5 many distribution functions F were either shown to have a density f with respect to Lebesgue measure or were assumed to have one, but such F 's were never intrinsically characterized, as they will be in this section.

Derivatives of Integrals

The first step is to show that a nondecreasing function has a derivative almost everywhere. This requires two preliminary results. Let λ denote Lebesgue measure.

Lemma 1. *Let A be a bounded linear Borel set, and let \mathcal{J} be a collection of open intervals covering A . Then \mathcal{J} contains a finite, disjoint subcollection I_1, \dots, I_k for which $\sum_{i=1}^k \lambda(I_i) \geq \lambda(A)/6$.*

PROOF. By regularity (Theorem 12.3) A contains a compact subset K satisfying $\lambda(K) \geq \lambda(A)/2$. Choose in \mathcal{J} a finite subcollection \mathcal{J}_0 covering K . Let I_1 be an interval in \mathcal{J}_0 of maximal length; discard from \mathcal{J}_0 the interval I_1 and all the others that intersect I_1 . Among the intervals remaining in \mathcal{J}_0 , let I_2 be one of maximal length; discard I_2 and all intervals that intersect it. Continue this way until \mathcal{J}_0 is exhausted. The I_i are disjoint. Let J_i be the interval with the same midpoint as I_i and three times the length. If I is an interval in \mathcal{J}_0 that is cast out because it meets I_i , then $I \subset J_i$. Thus each discarded interval is contained in one of the J_i , and so the J_i cover K . Hence $\sum \lambda(I_i) = \sum \lambda(J_i)/3 \geq \lambda(K)/3 \geq \lambda(A)/6$. ■

If

$$(31.4) \quad \Delta: a = a_0 < a_1 < \cdots < a_k = b$$

is a partition of an interval $[a, b]$ and F is a function over $[a, b]$, let

$$(31.5) \quad \|F\|_{\Delta} = \sum_{i=1}^k |F(a_i) - F(a_{i-1})|.$$

Lemma 2. *Consider a partition (31.4) and a nonnegative θ . If*

$$(31.6) \quad F(a) \leq F(b),$$

and if

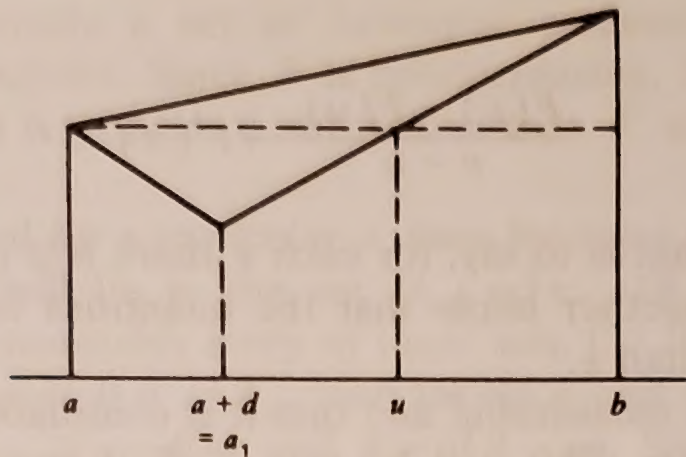
$$(31.7) \quad \frac{F(a_i) - F(a_{i-1})}{a_i - a_{i-1}} \leq -\theta$$

for a set of intervals $[a_{i-1}, a_i]$ of total length d , then

$$\|F\|_{\Delta} \geq |F(b) - F(a)| + 2\theta d.$$

This also holds if the inequalities in (31.6) and (31.7) are reversed and $-\theta$ is replaced by θ in the latter.

PROOF. The figure shows the case where $k = 2$ and the left-hand interval satisfies (31.7). Here F falls at least θd over $[a, a + d]$, rises the same amount over $[a + d, u]$, and then rises $F(b) - F(a)$ over $[u, b]$.



For the general case, let Σ' denote summation over those i satisfying (31.7) and let Σ'' denote summation over the remaining i ($1 \leq i \leq k$). Then

$$\begin{aligned} \|F\|_{\Delta} &= \Sigma' (F(a_{i-1}) - F(a_i)) + \Sigma'' |F(a_i) - F(a_{i-1})| \\ &\geq \Sigma' (F(a_{i-1}) - F(a_i)) + \left| \Sigma'' (F(a_i) - F(a_{i-1})) \right| \\ &= \Sigma' (F(a_{i-1}) - F(a_i)) + |(F(b) - F(a)) + \Sigma' (F(a_{i-1}) - F(a_i))|. \end{aligned}$$

As all the differences in this last expression are nonnegative, the absolute-value bars can be suppressed; therefore,

$$\begin{aligned} \|F\|_{\Delta} &\geq F(b) - F(a) + 2 \Sigma' (F(a_{i-1}) - F(a_i)) \\ &\geq F(b) - F(a) + 2\theta \Sigma' (a_i - a_{i-1}). \end{aligned}$$

A function F has at each x four *derivates*, the upper and lower right derivatives

$$D^F(x) = \limsup_{h \downarrow 0} \frac{F(x+h) - F(x)}{h},$$

$$D_F(x) = \liminf_{h \downarrow 0} \frac{F(x+h) - F(x)}{h},$$

and the upper and lower left derivatives

$${}^FD(x) = \limsup_{h \downarrow 0} \frac{F(x) - F(x-h)}{h},$$

$${}_FD(x) = \liminf_{h \downarrow 0} \frac{F(x) - F(x-h)}{h}.$$

There is a derivative at x if and only if these four quantities have a common value. Suppose that F has finite derivative $F'(x)$ at x . If $u \leq x \leq v$, then

$$\left| \frac{F(v) - F(u)}{v - u} - F'(x) \right| \leq \frac{v - x}{v - u} \left| \frac{F(v) - F(x)}{v - x} - F'(x) \right| + \frac{x - u}{v - u} \left| \frac{F(x) - F(u)}{x - u} - F'(x) \right|.$$

Therefore,

$$(31.8) \quad \frac{F(v) - F(u)}{v - u} \rightarrow F'(x)$$

as $u \uparrow x$ and $v \downarrow x$; that is to say, for each ϵ there is a δ such that $u \leq x \leq v$ and $0 < v - u < \delta$ together imply that the quantities on either side of the arrow differ by less than ϵ .

Suppose that F is measurable and that it is continuous except possibly at countably many points. This will be true if F is nondecreasing or is the difference of two nondecreasing functions. Let M be a countable, dense set containing all the discontinuity points of F ; let $r_n(x)$ be the smallest number of the form k/n exceeding x . Then

$$D^F(x) = \lim_{n \rightarrow \infty} \sup_{\substack{x < y < r_n(x) \\ y \in M}} \frac{F(y) - F(x)}{y - x};$$

the function inside the limit is measurable because the x -set where it exceeds α is

$$\bigcup_{y \in M} [x: x < y < r_n(x), F(y) - F(x) > \alpha(y - x)].$$

Thus $D^F(x)$ is measurable, as are the other three derivatives. This does not exclude infinite values. The set where the four derivatives have a common finite value F' is therefore a Borel set. In the following theorem, set $F' = 0$ (say) outside this set; F' is then a Borel function.

Theorem 31.2. *A nondecreasing function F is differentiable almost everywhere, the derivative F' is nonnegative, and*

$$(31.9) \quad \int_a^b F'(t) dt \leq F(b) - F(a)$$

for all a and b .

This and the following theorems can also be formulated for functions over an interval.

PROOF. If it can be shown that

$$(31.10) \quad D^F(x) \leq_F D(x)$$

except on a set of Lebesgue measure 0, then by the same result applied to $G(x) = -F(-x)$ it will follow that $^F D(x) = D^G(-x) \leq_G D(-x) = D_F(x)$ almost everywhere. This will imply that $D_F(x) \leq D^F(x) \leq_F D(x) \leq ^F D(x) \leq D_F(x)$ almost everywhere, since the first and third of these inequalities are obvious, and so, outside a set of Lebesgue measure 0, F will have a derivative, possibly infinite. Since F is nondecreasing, F' must be nonnegative, and once (31.9) is proved, it will follow that F' is finite almost everywhere.

If (31.10) is violated for a particular x , then for some pair α, β of rationals satisfying $\alpha < \beta$, x will lie in the set $A_{\alpha\beta} = [x: {}_F D(x) < \alpha < \beta < D^F(x)]$. Since there are only countably many of these sets, (31.10) will hold outside a set of Lebesgue measure 0 if $\lambda(A_{\alpha\beta}) = 0$ for all α and β .

Put $G(x) = F(x) - \frac{1}{2}(\alpha + \beta)x$ and $\theta = \frac{1}{2}(\beta - \alpha)$. Since differentiation is linear, $A_{\alpha\beta} = B_\theta = [x: {}_G D(x) < -\theta < \theta < D^G(x)]$. Since F and G have only countably many discontinuities, it suffices to prove that $\lambda(C_\theta) = 0$, where C_θ is the set of points in B_θ that are continuity points of G . Consider an interval (a, b) , and suppose for the moment that $G(a) \leq G(b)$. For each x in C_θ satisfying $a < x < b$, from ${}_G D(x) < -\theta$ it follows that there exists an open interval (a_x, b_x) for which $x \in (a_x, b_x) \subset (a, b)$ and

$$(31.11) \quad \frac{G(b_x) - G(a_x)}{b_x - a_x} < -\theta.$$

There exists by Lemma 1 a finite, disjoint collection (a_{x_i}, b_{x_i}) of these intervals of total length $\sum(b_{x_i} - a_{x_i}) \geq \lambda((a, b) \cap C_\theta)/6$. Let Δ be the partition (31.4) of $[a, b]$ with the points a_{x_i} and b_{x_i} in the role of the a_1, \dots, a_{k-1} . By Lemma 2,

$$(31.12) \quad \|G\|_\Delta \geq |G(b) - G(a)| + \frac{1}{3}\theta\lambda((a, b) \cap C_\theta).$$

If instead of $G(a) \leq G(b)$ the reverse inequality holds, choose a_x and b_x so that the ratio in (31.11) exceeds θ , which is possible because $D^G(x) > \theta$ for $x \in C_\theta$. Again (31.12) follows.

In each interval $[a, b]$ there is thus a partition (31.4) satisfying (31.12). Apply this to each interval $[a_{i-1}, a_i]$ in the partition. This gives a partition Δ_1 that refines Δ , and adding the corresponding inequalities (31.12) leads to

$$\|G\|_{\Delta_1} \geq \|G\|_\Delta + \frac{1}{3}\theta\lambda((a, b) \cap C_\theta).$$

Continuing leads to a sequence of successively finer partitions Δ_n such that

$$(31.13) \quad \|G\|_{\Delta_n} \geq n \frac{\theta}{3} \lambda((a, b) \cap C_\theta).$$

Now $\|G\|_\Delta$ is bounded by $|F(b) - F(a)| + \frac{1}{2}|\alpha + \beta|(b - a)$ because F is monotonic. Thus (31.13) is impossible unless $\lambda((a, b) \cap C_\theta) = 0$. Since (a, b) can be any interval, $\lambda(C_\theta) = 0$. This proves (31.10) and establishes the differentiability of F almost everywhere.

It remains to prove (31.9). Let

$$(31.14) \quad f_n(x) = \frac{F(x + n^{-1}) - F(x)}{n^{-1}}.$$

Now f_n is nonnegative, and by what has been shown, $f_n(x) \rightarrow F'(x)$ except on a set of Lebesgue measure 0. By Fatou's lemma and the fact that F is nondecreasing,

$$\begin{aligned} \int_a^b F'(x) dx &\leq \liminf_n \int_a^b f_n(x) dx \\ &= \liminf_n \left[n \int_b^{b+n^{-1}} F(x) dx - n \int_a^{a+n^{-1}} F(x) dx \right] \\ &\leq \liminf_n [F(b + n^{-1}) - F(a)] = F(b+) - F(a). \end{aligned}$$

Replacing b by $b - \epsilon$ and letting $\epsilon \rightarrow 0$ gives (31.9). ■

Theorem 31.3. *If f is nonnegative and integrable, and if $F(x) = \int_{-\infty}^x f(t) dt$, then $F'(x) = f(x)$ except on a set of Lebesgue measure 0.*

Since f is nonnegative, F is nondecreasing and hence by Theorem 31.2 is differentiable almost everywhere. The problem is to show that the derivative F' coincides with f almost everywhere.

PROOF FOR BOUNDED f . Suppose first that f is bounded by M . Define f_n by (31.14). Then $f_n(x) = n \int_x^{x+n^{-1}} f(t) dt$ is bounded by M and converges almost everywhere to $F'(x)$, so that the bounded convergence theorem gives

$$\begin{aligned} \int_a^b F'(x) dx &= \lim_n \int_a^b f_n(x) dx \\ &= \lim_n \left[n \int_b^{b+n^{-1}} F(x) dx - n \int_a^{a+n^{-1}} F(x) dx \right]. \end{aligned}$$

Since F is continuous (see (31.3)), this last limit is $F(b) - F(a) = \int_a^b f(x) dx$.

Thus $\int_A F'(x) dx = \int_A f(x) dx$ for bounded intervals $A = (a, b]$. Since these form a π -system, it follows (Theorem 16.10(iii)) that $F' = f$ almost everywhere. ■

PROOF FOR INTEGRABLE f . Apply the result for bounded functions to f truncated at n : If $h_n(x)$ is $f(x)$ or n as $f(x) \leq n$ or $f(x) > n$, then $H_n(x) = \int_{-\infty}^x h_n(t) dt$ differentiates almost everywhere to $h_n(x)$ by the case already treated. Now $F(x) = H_n(x) + \int_{-\infty}^x (f(t) - h_n(t)) dt$; the integral here is nondecreasing because the integrand is nonnegative, and it follows by Theorem 31.2 that it has almost everywhere a nonnegative derivative. Since differentiation is linear, $F'(x) \geq H'_n(x) = h_n(x)$ almost everywhere. As n was arbitrary, $F'(x) \geq f(x)$ almost everywhere, and so $\int_a^b F'(x) dx \geq \int_a^b f(x) dx = F(b) - F(a)$. But the reverse inequality is a consequence of (31.9). Therefore, $\int_a^b (F'(x) - f(x)) dx = 0$, and as before $F' = f$ except on a set of Lebesgue measure 0. ■

Singular Functions

If $f(x)$ is nonnegative and integrable, differentiating its indefinite integral $\int_{-\infty}^x f(t) dt$ leads back to $f(x)$ except perhaps on a set of Lebesgue measure 0. That is the content of Theorem 31.3. The converse question is this: If $F(x)$ is nondecreasing and hence has almost everywhere a derivative $F'(x)$, does integrating $F'(x)$ lead back to $F(x)$? As stated before, the answer turns out to be no even if $F(x)$ is assumed continuous:

Example 31.1. Let X_1, X_2, \dots be independent, identically distributed random variables such that $P[X_n = 0] = p_0$ and $P[X_n = 1] = p_1 = 1 - p_0$, and let $X = \sum_{n=1}^{\infty} X_n 2^{-n}$. Let $F(x) = P[X \leq x]$ be the distribution function of X . For an arbitrary sequence u_1, u_2, \dots of 0's and 1's, $P[X_n = u_n, n = 1, 2, \dots] = \lim_n p_{u_1} \cdots p_{u_n} = 0$; since x can have at most two dyadic expansions $x = \sum_n u_n 2^{-n}$, $P[X = x] = 0$. Thus F is everywhere continuous. Of course, $F(0) = 0$ and $F(1) = 1$. For $0 \leq k < 2^n$, $k 2^{-n}$ has the form $\sum_{i=1}^n u_i 2^{-i}$ for some n -tuple (u_1, \dots, u_n) of 0's and 1's. Since F is continuous,

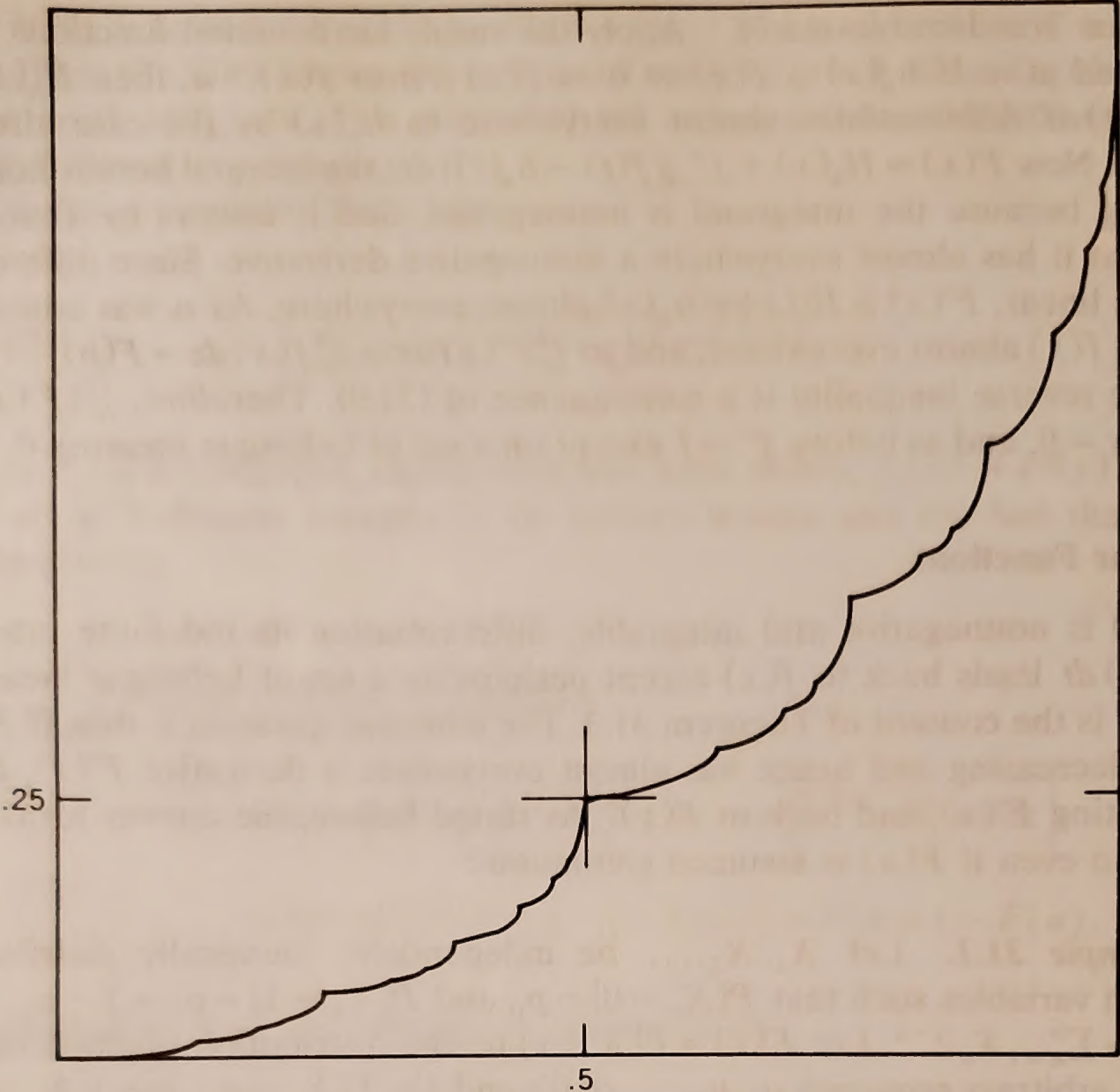
$$(31.15) \quad F\left(\frac{k+1}{2^n}\right) - F\left(\frac{k}{2^n}\right) = P\left[\frac{k}{2^n} < X < \frac{k+1}{2^n}\right] \\ = P[X_i = u_i, i \leq n] = p_{u_1} \cdots p_{u_n}.$$

This shows that F is strictly increasing over the unit interval.

If $p_0 = p_1 = \frac{1}{2}$, the right side of (31.15) is 2^{-n} , and a passage to the limit shows that $F(x) = x$ for $0 \leq x \leq 1$. Assume, however, that $p_0 \neq p_1$. It will be shown that $F'(x) = 0$ except on a set of Lebesgue measure 0 in this case. Obviously the derivative is 0 outside the unit interval, and by Theorem 31.2 it exists almost everywhere inside it. Suppose then that $0 < x < 1$ and that F has a derivative $F'(x)$ at x . It will be shown that $F'(x) = 0$.

For each n choose k_n so that x lies in the interval $I_n = (k_n 2^{-n}, (k_n + 1) 2^{-n}]$; I_n is that dyadic interval of rank n that contains x . By (31.8),

$$\frac{P[X \in I_n]}{2^{-n}} = \frac{F((k_n + 1) 2^{-n}) - F(k_n 2^{-n})}{2^{-n}} \rightarrow F'(x).$$



Graph of $F(x)$ for $p_0 = .25, p_1 = .75$. Because of the recursion (31.17), the part of the graph over $[0, .5]$ and the part over $[.5, 1]$ are identical, apart from changes in scale, with the whole graph. Each segment of the curve therefore contains scaled copies of the whole; the extreme irregularity this implies is obscured by the fact that the accuracy is only to within the width of the printed line.

If $F'(x)$ is distinct from 0, the ratio of two successive terms here must go to 1, so that

(31.16)
$$\frac{P[X \in I_{n+1}]}{P[X \in I_n]} \rightarrow \frac{1}{2}.$$

If I_n consists of the reals with nonterminating base-2 expansions beginning with the digits u_1, \dots, u_n , then $P[X \in I_n] = p_{u_1} \cdots p_{u_n}$ by (31.15). But I_{n+1} must for some u_{n+1} consist of the reals beginning u_1, \dots, u_n, u_{n+1} (u_{n+1} is 1 or 0 according as x lies to the right of the midpoint of I_n or not). Thus $P[X \in I_{n+1}]/P[X \in I_n] = p_{u_{n+1}}$ is either p_0 or p_1 , and (31.16) is possible only if $p_0 = p_1$, which was excluded by hypothesis.

Thus F is continuous and strictly increasing over $[0, 1]$, but $F'(x) = 0$ except on a set of Lebesgue measure 0. For $0 \leq x \leq \frac{1}{2}$ independence gives

$F(x) = P[X_1 = 0, \sum_{n=2}^{\infty} X_n 2^{-n+1} \leq 2x] = p_0 F(2x)$. Similarly, $F(x) - p_0 = p_1 F(2x - 1)$ for $\frac{1}{2} \leq x \leq 1$. Thus

$$(31.17) \quad F(x) = \begin{cases} p_0 F(2x) & \text{if } 0 \leq x \leq \frac{1}{2}, \\ p_0 + p_1 F(2x - 1) & \text{if } \frac{1}{2} \leq x \leq 1. \end{cases}$$

In Section 7, $F(x)$ (there denoted $Q(x)$) entered as the probability of success at bold play; see (7.30) and (7.33). ■

A function is *singular* if it has derivative 0 except on a set of Lebesgue measure 0. Of course, a step function constant over intervals is singular. What is remarkable (indeed, singular) about the function in the preceding example is that it is continuous and strictly increasing but nonetheless has derivative 0 except on a set of Lebesgue measure 0. Note that there is strict inequality in (31.9) for this F .

Further properties of nondecreasing functions can be discovered through a study of the measures they generate. Assume from now on that F is nondecreasing, that F is continuous from the right (this is only a normalization), and that $0 = \lim_{x \rightarrow -\infty} F(x) \leq \lim_{x \rightarrow +\infty} F(x) = m < \infty$. Call such an F a *distribution function*, even though m need not be 1. By Theorem 12.4 there exists a unique measure μ on the Borel sets of the line for which

$$(31.18) \quad \mu(a, b] = F(b) - F(a).$$

Of course, $\mu(R^1) = m$ is finite.

The larger F' is, the larger μ is:

Theorem 31.4. *Suppose that F and μ are related by (31.18) and that $F'(x)$ exists throughout a Borel set A .*

- (i) *If $F'(x) \leq \alpha$ for $x \in A$, then $\mu(A) \leq \alpha \lambda(A)$.*
- (ii) *If $F'(x) \geq \alpha$ for $x \in A$, then $\mu(A) \geq \alpha \lambda(A)$.*

PROOF. It is no restriction to assume A bounded. Fix ϵ for the moment. Let E be a countable, dense set, and let $A_n = \bigcap (A \cap I)$, where the intersection extends over the intervals $I = (u, v]$ for which $u, v \in E$, $0 < \lambda(I) < n^{-1}$, and

$$(31.19) \quad \mu(I) < (\alpha + \epsilon) \lambda(I).$$

Then A_n is a Borel set and (see (31.8)) $A_n \uparrow A$ under the hypothesis of (i). By Theorem 11.4 there exist disjoint intervals I_{nk} (open on the left, closed on

the right) such that $A_n \subset \bigcup_k I_{nk}$ and

$$(31.20) \quad \sum_k \lambda(I_{nk}) < \lambda(A_n) + \epsilon.$$

It is no restriction to assume that each I_{nk} has endpoints in E , meets A_n , and satisfies $\lambda(I_{nk}) < n^{-1}$. Then (31.19) applies to each I_{nk} , and hence

$$\mu(A_n) \leq \sum_k \mu(I_{nk}) \leq (\alpha + \epsilon) \sum_k \lambda(I_{nk}) \leq (\alpha + \epsilon)(\lambda(A_n) + \epsilon).$$

In the extreme terms here let $n \rightarrow \infty$ and then $\epsilon \rightarrow 0$; (i) follows.

To prove (ii), let the countable, dense set E contain all the discontinuity points of F , and use the same argument with $\mu(I) \geq (\alpha - \epsilon)\lambda(I)$ in place of (31.19) and $\sum_k \mu(I_{nk}) < \mu(A_n) + \epsilon$ in place of (31.20). Since E contains all the discontinuity points of F , it is again no restriction to assume that each I_{nk} has endpoints in E , meets A_n , and satisfies $\lambda(I_{nk}) < n^{-1}$. It follows that

$$\mu(A_n) + \epsilon > \sum_k \mu(I_{nk}) \geq (\alpha - \epsilon) \sum_k \lambda(I_{nk}) \geq (\alpha - \epsilon)\lambda(A_n).$$

Again let $n \rightarrow \infty$ and then $\epsilon \rightarrow 0$. ■

The measures μ and λ have *disjoint supports* if there exist Borel sets S_μ and S_λ such that

$$(31.21) \quad \begin{cases} \mu(R^1 - S_\mu) = 0, & \lambda(R^1 - S_\lambda) = 0, \\ S_\mu \cap S_\lambda = 0. \end{cases}$$

Theorem 31.5. *Suppose that F and μ are related by (31.18). A necessary and sufficient condition for μ and λ to have disjoint supports is that $F'(x) = 0$ except on a set of Lebesgue measure 0.*

PROOF. By Theorem 31.4, $\mu[x: |x| \leq a, F'(x) \leq \epsilon] \leq 2a\epsilon$, and so (let $\epsilon \rightarrow 0$ and then $a \rightarrow \infty$) $\mu[x: F'(x) = 0] = 0$. If $F'(x) = 0$ outside a set of Lebesgue measure 0, then $S_\lambda = [x: F'(x) = 0]$ and $S_\mu = R^1 - S_\lambda$ satisfy (31.21).

Suppose that there exist S_μ and S_λ satisfying (31.21). By the other half of Theorem 31.4, $\epsilon\lambda[x: F'(x) \geq \epsilon] = \epsilon\lambda[x: x \in S_\lambda, F'(x) \geq \epsilon] \leq \mu(S_\lambda) = 0$, and so (let $\epsilon \rightarrow 0$) $F'(x) = 0$ except on a set of Lebesgue measure 0. ■

Example 31.2. Suppose that μ is discrete, consisting of a mass m_k at each of countably many points x_k . Then $F(x) = \sum m_k$, the sum extending over the k for which $x_k \leq x$. Certainly, μ and λ have disjoint supports, and so F' must vanish except on a set of Lebesgue measure 0. This is directly obvious if the x_k have no limit points, but not, for example, if they are dense. ■

Example 31.3. Consider again the distribution function F in Example 31.1. Here $\mu(A) = P[X \in A]$. Since F is singular, μ and λ have disjoint supports. This fact has an interesting direct probabilistic proof.

For x in the unit interval, let $d_1(x), d_2(x), \dots$ be the digits in its nonterminating dyadic expansion, as in Section 1. If $(k2^{-n}, (k+1)2^{-n}]$ is the dyadic interval of rank n consisting of the reals whose expansions begin with the digits u_1, \dots, u_n , then, by (31.15),

$$(31.22) \quad \mu\left(\frac{k}{2^n}, \frac{k+1}{2^n}\right] = \mu[x: d_i(x) = u_i, i \leq n] = p_{u_1} \cdots p_{u_n}.$$

If the unit interval is regarded as a probability space under the measure μ , then the $d_i(x)$ become random variables, and (31.22) says that these random variables are independent and identically distributed and $\mu[x: d_i(x) = 0] = p_0$, $\mu[x: d_i(x) = 1] = p_1$.

Since these random variables have expected value p_1 , the strong law of large numbers implies that their averages go to p_1 with probability 1:

$$(31.23) \quad \mu\left[x \in (0, 1]: \lim_n \frac{1}{n} \sum_{i=1}^n d_i(x) = p_1\right] = 1.$$

On the other hand, by the normal number theorem,

$$(31.24) \quad \lambda\left[x \in (0, 1]: \lim_n \frac{1}{n} \sum_{i=1}^n d_i(x) = \frac{1}{2}\right] = 1.$$

(Of course, (31.24) is just (31.23) for the special case $p_0 = p_1 = \frac{1}{2}$; in this case μ and λ coincide in the unit interval.) If $p_1 \neq \frac{1}{2}$, the sets in (31.23) and (31.24) are disjoint, so that μ and λ do have disjoint supports.

It was shown in Example 31.1 that if $F'(x)$ exists at all ($0 < x < 1$), then it is 0. By part (i) of Theorem 31.4 the set where $F'(x)$ fails to exist therefore has μ -measure 1; in particular, this set is uncountable. ■

In the singular case, according to Theorem 31.5, F' vanishes on a support of λ . It is natural to ask for the size of F' on a support of μ . If B is the x -set where F has a finite derivative, and if (31.21) holds, then by Theorem 31.4, $\mu[x \in B: F'(x) \leq n] = \mu[x \in B \cap S_\mu: F'(x) \leq n] \leq n\lambda(S_\mu) = 0$, and hence $\mu(B) = 0$. The next theorem goes further.

Theorem 31.6. Suppose that F and μ are related by (31.18) and that μ and λ have disjoint supports. Then, except for x in a set of μ -measure 0, ${}_F D(x) = \infty$.

If μ has finite support, then clearly ${}_F D(x) = \infty$ if $\mu\{x\} > 0$, while $D_F(x) = 0$ for all x . Since F is continuous from the right, ${}_F D$ and D_F play different roles.[†]

PROOF. Let A_n be the set where ${}_F D(x) < n$. The problem is to prove that $\mu(A_n) = 0$, and by (31.21) it is enough to prove that $\mu(A_n \cap S_\mu) = 0$. Further, by Theorem 12.3 it is enough to prove that $\mu(K) = 0$ if K is a compact subset of $A_n \cap S_\mu$.

Fix ϵ . Since $\lambda(K) = 0$, there is an open G such that $K \subset G$ and $\lambda(G) < \epsilon$. If $x \in K$, then $x \in A_n$, and by the definition of ${}_F D$ and the right-continuity of F , there is an open interval I_x for which $x \in I_x \subset G$ and $\mu(I_x) < n\lambda(I_x)$. By compactness, K has a finite subcover I_{x_1}, \dots, I_{x_k} . If some three of these have a nonempty intersection, one of them must be contained in the union of the other two. Such superfluous intervals can be removed from the subcover, and it is therefore possible to assume that no point of K lies in more than two of the I_{x_i} . But then

$$\begin{aligned} \mu(K) &\leq \mu\left(\bigcup_i I_{x_i}\right) \leq \sum_i \mu(I_{x_i}) \leq n \sum_i \lambda(I_{x_i}) \\ &\leq 2n\lambda\left(\bigcup_i I_{x_i}\right) \leq 2n\lambda(G) \leq 2n\epsilon. \end{aligned}$$

Since ϵ was arbitrary, $\lambda(K) = 0$, as required. ■

Example 31.4. Restrict the F of Examples 31.1 and 31.3 to $(0, 1)$, and let g be the inverse. Thus F and g are continuous, strictly increasing mappings of $(0, 1)$ onto itself. If $A = [x \in (0, 1): F'(x) = 0]$, then $\lambda(A) = 1$, as shown in the examples, while $\mu(A) = 0$. Let H be a set in $(0, 1)$ that is not a Lebesgue set. Since $H - A$ is contained in a set of Lebesgue measure 0, it is a Lebesgue set; hence $H_0 = H \cap A$ is not a Lebesgue set, since otherwise $H = H_0 \cup (H - A)$ would be a Lebesgue set. If $B = (0, x]$, then $\lambda g^{-1}(B) = \lambda(0, F(x)] = F(x) = \mu(B)$, and it follows that $\lambda g^{-1}(B) = \mu(B)$ for all Borel sets B . Since $g^{-1}H_0$ is a subset of $g^{-1}A$ and $\lambda(g^{-1}A) = \mu(A) = 0$, $g^{-1}H_0$ is a Lebesgue set. On the other hand, if $g^{-1}H_0$ were a Borel set, $H_0 = F^{-1}(g^{-1}H_0)$ would also be a Borel set. Thus $g^{-1}H_0$ provides an example of a Lebesgue set that is not a Borel set.[‡] ■

Integrals of Derivatives

Return now to the problem of extending part (ii) of Theorem 31.1, to the problem of characterizing those distribution functions F for which F' integrates back to F :

$$(31.25) \quad F(x) = \int_{-\infty}^x F'(t) dt.$$

[†]See Problem 31.8.

[‡]For a different argument, see Problem 3.14.

The first step is easy: If (31.25) holds, then F has the form

$$(31.26) \quad F(x) = \int_{-\infty}^x f(t) dt$$

for a nonnegative, integrable f (a density), namely $f = F'$. On the other hand, (31.26) implies by Theorem 31.3 that $F' = f$ outside a set of Lebesgue measure 0, whence (31.25) follows. Thus (31.25) holds if and only if F has the form (31.26) for some f , and the problem is to characterize functions of this form. The function of Example 31.1 is not among them.

As observed earlier (see (31.3)), an F of the form (31.26) with f integrable is continuous. It has a still stronger property: For each ϵ there exists a δ such that

$$(31.27) \quad \int_A f(x) dx < \epsilon \quad \text{if } \lambda(A) < \delta.$$

Indeed, if $A_n = [x: f(x) > n]$, then $A_n \downarrow \emptyset$, and since f is integrable, the dominated convergence theorem implies that $\int_{A_n} f(x) dx < \epsilon/2$ for large n . Fix such an n and take $\delta = \epsilon/2n$. If $\lambda(A) < \delta$, then $\int_A f(x) dx \leq \int_{A-A_n} f(x) dx + \int_{A_n} f(x) dx \leq n\lambda(A) + \epsilon/2 < \epsilon$.

If F is given by (31.26), then $F(b) - F(a) = \int_a^b f(x) dx$, and (31.27) has this consequence: For every ϵ there exists a δ such that for each finite collection $[a_i, b_i]$, $i = 1, \dots, k$, of nonoverlapping[†] intervals,

$$(31.28) \quad \sum_{i=1}^k |F(b_i) - F(a_i)| < \epsilon \quad \text{if } \sum_{i=1}^k (b_i - a_i) < \delta.$$

A function F with this property is said to be *absolutely continuous*.[‡] A function of the form (31.26) (f integrable) is thus absolutely continuous.

A continuous distribution function is uniformly continuous, and so for every ϵ there is a δ such that the implication in (31.28) holds provided that $k = 1$. The definition of absolute continuity requires this to hold whatever k may be, which puts severe restrictions on F . Absolute continuity of F can be characterized in terms of the measure μ :

Theorem 31.7. *Suppose that F and μ are related by (31.18). Then F is absolutely continuous in the sense of (31.28) if and only if $\mu(A) = 0$ for every A for which $\lambda(A) = 0$.*

[†]Intervals are nonoverlapping if their interiors are disjoint. In this definition it is immaterial whether the intervals are regarded as closed or open or half-open, since this has no effect on (31.28).

[‡]The definition applies to all functions, not just to distribution functions. If F is a distribution function as in the present discussion, the absolute-value bars in (31.28) are unnecessary.

PROOF. Suppose that F is absolutely continuous and that $\lambda(A) = 0$. Given ϵ , choose δ so that (31.28) holds. There exists a countable disjoint union $B = \bigcup_k I_k$ of intervals such that $A \subset B$ and $\lambda(B) < \delta$. By (31.28) it follows that $\mu(\bigcup_{k=1}^n I_k) \leq \epsilon$ for each n and hence that $\mu(A) \leq \mu(B) \leq \epsilon$. Since ϵ was arbitrary, $\mu(A) = 0$.

If F is not absolutely continuous, then there exists an ϵ such that for every δ some finite disjoint union A of intervals satisfies $\lambda(A) < \delta$ and $\mu(A) \geq \epsilon$. Choose A_n so that $\lambda(A_n) < n^{-2}$ and $\mu(A_n) \geq \epsilon$. Then $\lambda(\limsup_n A_n) = 0$ by the first Borel–Cantelli lemma (Theorem 4.3, the proof of which does not require P to be a probability measure or even finite). On the other hand, $\mu(\limsup_n A_n) \geq \epsilon > 0$ by Theorem 4.1 (the proof of which applies because μ is assumed finite). ■

This result leads to a characterization of indefinite integrals.

Theorem 31.8. *A distribution function $F(x)$ has the form $\int_{-\infty}^x f(t) dt$ for an integrable f if and only if it is absolutely continuous in the sense of (31.28).*

PROOF. That an F of the form (31.26) is absolutely continuous was proved in the argument leading to the definition (31.28). For another proof, apply Theorem 31.7: if F has this form, then $\lambda(A) = 0$ implies that $\mu(A) = \int_A f(t) dt = 0$.

To go the other way, define for any distribution function F

$$(31.29) \quad F_{ac}(x) = \int_{-\infty}^x F'(t) dt$$

and

$$(31.30) \quad F_s(x) = F(x) - F_{ac}(x).$$

Then F_s is right-continuous, and by (31.9) it is both nonnegative and nondecreasing. Since F_{ac} comes from a density, it is absolutely continuous. By Theorem 31.3, $F'_{ac} = F'$ and hence $F'_s = 0$ except on a set of Lebesgue measure 0. Thus F has a decomposition

$$(31.31) \quad F(x) = F_{ac}(x) + F_s(x),$$

where F_{ac} has a density and hence is absolutely continuous and F_s is singular. This is called the *Lebesgue decomposition*.

Suppose that F is absolutely continuous. Then F_s of (31.30) must, as the difference of absolutely continuous functions, be absolutely continuous itself. If it can be shown that F_s is identically 0, it will follow that $F = F_{ac}$ has the required form. It thus suffices to show that a distribution function that is both absolutely continuous and singular must vanish.

If a distribution function F is singular, then by Theorem 31.5 there are disjoint supports S_μ and S_λ . But if F is also absolutely continuous, then from $\lambda(S_\mu) = 0$ it follows by Theorem 31.7 that $\mu(S_\mu) = 0$. But then $\mu(R^1) = 0$, and so $F(x) \equiv 0$. ■

This theorem identifies the distribution functions that are integrals of their derivatives as the absolutely continuous functions. Theorem 31.7, on the other hand, characterizes absolute continuity in a way that extends to spaces Ω without the geometric structure of the line necessary to a treatment involving distribution functions and ordinary derivatives.[†] The extension is studied in Section 32.

Functions of Bounded Variation

The remainder of this section briefly sketches the extension of the preceding theory to functions that are not monotone. The results are for simplicity given only for a finite interval $[a, b]$ and for functions F on $[a, b]$ satisfying $F(a) = 0$.

If $F(x) = \int_a^x f(t) dt$ is an indefinite integral, where f is integrable but not necessarily nonnegative, then $F(x) = \int_a^x f^+(t) dt - \int_a^x f^-(t) dt$ exhibits F as the difference of two nondecreasing functions. The problem of characterizing indefinite integrals thus leads to the preliminary problem of characterizing functions representable as a difference of nondecreasing functions.

Now F is said to be of *bounded variation* over $[a, b]$ if $\sup_\Delta \|F\|_\Delta$ is finite, where $\|F\|_\Delta$ is defined by (31.5) and Δ ranges over all partitions (31.4) of $[a, b]$. Clearly, a difference of nondecreasing functions is of bounded variation. But the converse holds as well: For every finite collection Γ of nonoverlapping intervals $[x_i, y_i]$ in $[a, b]$, put

$$P_\Gamma = \sum (F(y_i) - F(x_i))^+, \quad N_\Gamma = \sum (F(y_i) - F(x_i))^-.$$

Now define

$$P(x) = \sup_\Gamma P_\Gamma, \quad N(x) = \sup_\Gamma N_\Gamma,$$

where the suprema extend over partitions Γ of $[a, x]$. If F is of bounded variation, then $P(x)$ and $N(x)$ are finite. For each such Γ , $P_\Gamma = N_\Gamma + F(x)$. This gives the inequalities

$$P_\Gamma \leq N(x) + F(x), \quad P(x) \geq N_\Gamma + F(x),$$

which in turn lead to the inequalities

$$P(x) \leq N(x) + F(x), \quad P(x) \geq N(x) + F(x).$$

Thus

$$(31.32) \quad F(x) = P(x) - N(x)$$

gives the required representation: *A function is the difference of two nondecreasing functions if and only if it is of bounded variation.*

[†]Theorems 31.3 and 31.8 do have geometric analogues in R^k ; see RUDIN₂, Chapter 8.

If $T_\Gamma = P_\Gamma + N_\Gamma$, then $T_\Gamma = \sum |F(y_i) - F(x_i)|$. According to the definition (31.28), F is absolutely continuous if for every ϵ there exists a δ such that $T_\Gamma < \epsilon$ whenever the intervals in the collection Γ have total length less than δ . If F is absolutely continuous, take the δ corresponding to an ϵ of 1 and decompose $[a, b]$ into a finite number, say n , of subintervals $[u_{j-1}, u_j]$ of lengths less than δ . Any partition Δ of $[a, b]$ can by the insertion of the u_j be split into n sets of intervals each of total length less than δ , and it follows[†] that $\|F\|_\Delta \leq n$. Therefore, an absolutely continuous function is necessarily of bounded variation.

An absolutely continuous F thus has a representation (31.32). It follows by the definitions that $P(y) - P(x)$ is at most $\sup_\Gamma T_\Gamma$, where Γ ranges over the partitions of $[x, y]$. If $[x_i, y_i]$ are nonoverlapping intervals, then $\sum (P(y_i) - P(x_i))$ is at most $\sup_\Gamma T_\Gamma$, where now Γ ranges over the collections of intervals that partition *each* of the $[x_i, y_i]$. Therefore, if F is absolutely continuous, there exists for each ϵ a δ such that $\sum (y_i - x_i) < \delta$ implies that $\sum (P(y_i) - P(x_i)) < \epsilon$. In other words, P is absolutely continuous. Similarly, N is absolutely continuous.

Thus an absolutely continuous F is the difference of two nondecreasing absolutely continuous functions. By Theorem 31.8, each of these is an indefinite integral, which implies that F is an indefinite integral as well: *For an F on $[a, b]$ satisfying $F(a) = 0$, absolute continuity is a necessary and sufficient condition for F to be an indefinite integral—to have the form $F(x) = \int_a^x f(t) dt$ for an integrable f .*

PROBLEMS

- 31.1.** Extend Examples 31.1 and 31.3: Let p_0, \dots, p_{r-1} be nonnegative numbers adding to 1, where $r \geq 2$; suppose there is no i such that $p_i = 1$. Let X_1, X_2, \dots be independent, identically distributed random variables such that $P[X_n = i] = p_i$, $0 \leq i < r$, and put $X = \sum_{n=1}^\infty X_n r^{-n}$. Let F be the distribution function of X . Show that F is continuous. Show that F is strictly increasing over the unit interval if and only if all the p_i are strictly positive. Show that $F(x) \equiv x$ for $0 \leq x \leq 1$ if $p_i \equiv r^{-1}$ and that otherwise F is singular; prove singularity by extending the arguments both of Example 31.1 and of Example 31.3. What is the analogue of (31.17)?
- 31.2.** \uparrow In Problem 31.1 take $r = 3$ and $p_0 = p_2 = \frac{1}{2}$, $p_1 = 0$. The corresponding F is called the *Cantor function*. The complement in $[0, 1]$ of the Cantor set (see Problems 1.5 and 3.16) consists of the middle third $(\frac{1}{3}, \frac{2}{3})$, the middle thirds $(\frac{1}{9}, \frac{2}{9})$ and $(\frac{7}{9}, \frac{8}{9})$, and so on. Show that F is $\frac{1}{2}$ on the first of these intervals, $\frac{1}{4}$ on the second, $\frac{3}{4}$ on the third, and so on. Show by direct argument that $F' = 0$ except on a set of Lebesgue measure 0.
- 31.3.** A real function f of a real variable is a *Lebesgue function* if $[x: f(x) \leq \alpha]$ is a Lebesgue set for each α .
- (a) Show that, if f_1 is a Borel function and f_2 is a Lebesgue function, then the composition $f_1 f_2$ is a Lebesgue function.
- (b) Show that there exists a Lebesgue function f_1 and a Lebesgue (even Borel, even continuous) function f_2 such that $f_1 f_2$ is not a Lebesgue function. *Hint:* Use Example 31.4.

[†]This uses the fact that $\|F\|_\Delta$ cannot decrease under passage to a finer partition.

- 31.4.** \uparrow An arbitrary function f on $(0, 1]$ can be represented as a composition of a Lebesgue function f_1 and a Borel function f_2 . For x in $(0, 1]$, let $d_n(x)$ be the n th digit in its nonterminating dyadic expansion, and define $f_2(x) = \sum_{n=1}^{\infty} 2d_n(x)/3^n$. Show that f_2 is increasing and that $f_2(0, 1]$ is contained in the Cantor set. Take $f_1(x)$ to be $f(f_2^{-1}(x))$ if $x \in f_2(0, 1]$ and 0 if $x \in (0, 1] - f_2(0, 1]$. Now show that $f = f_1 f_2$.
- 31.5.** Let r_1, r_2, \dots be an enumeration of the rationals in $(0, 1)$ and put $F(x) = \sum_{k: r_k \leq x} 2^{-k}$. Define φ by (14.5) and prove that it is continuous and singular.
- 31.6.** Suppose that μ and F are related by (31.18). If F is not absolutely continuous, then $\mu(A) > 0$ for some set A of Lebesgue measure 0. It is an interesting fact, however, that almost all translates of A must have μ -measure 0. From Fubini's theorem and the fact that λ is invariant under translation and reflection through 0, show that, if $\lambda(A) = 0$ and μ is σ -finite, then $\mu(A + x) = 0$ for x outside a set of Lebesgue measure 0.
- 31.7.** 17.4 31.6 \uparrow Show that F is absolutely continuous if and only if for each Borel set A , $\mu(A + x)$ is continuous in x .
- 31.8.** Let $F_*(x) = \lim_{\delta \rightarrow 0} \inf (F(v) - F(u))/(v - u)$, where the infimum extends over u and v such that $u < x < v$ and $v - u < \delta$. Define $F^*(x)$ as this limit with the infimum replaced by a supremum. Show that in Theorem 31.4, F' can be replaced by F^* in part (i) and by F_* in part (ii). Show that in Theorem 31.6, ${}_F D$ can be replaced by F_* (note that $F_*(x) \leq {}_F D(x)$).
- 31.9.** *Lebesgue's density theorem.* A point x is a *density point* of a Borel set A if $\lambda((u, v] \cap A)/(v - u) \rightarrow 1$ as $u \uparrow x$ and $v \downarrow x$. From Theorems 31.2 and 31.4 deduce that almost all points of A are density points. Similarly, $\lambda((u, v] \cap A)/(v - u) \rightarrow 0$ almost everywhere on A^c .
- 31.10.** Let $f: [a, b] \rightarrow R^k$ be an arc; $f(t) = (f_1(t), \dots, f_k(t))$. Show that the arc is rectifiable if and only if each f_i is of bounded variation over $[a, b]$.
- 31.11.** \uparrow Suppose that F is continuous and nondecreasing and that $F(0) = 0$, $F(1) = 1$. Then $f(x) = (x, F(x))$ defines an arc $f: [0, 1] \rightarrow R^2$. It is easy to see by monotonicity that the arc is rectifiable and that, in fact, its length satisfies $L(f) \leq 2$. It is also easy, given ϵ , to produce functions F for which $L(f) > 2 - \epsilon$. Show by the arguments in the proof of Theorem 31.4 that $L(f) = 2$ if F is singular.
- 31.12.** Suppose that the characteristic function of F satisfies $\limsup_{t \rightarrow \infty} |\varphi(t)| = 1$. Show that F is singular. Compare the lattice case (Problem 26.1). *Hint:* Use the Lebesgue decomposition and the Riemann–Lebesgue theorem.
- 31.13.** Suppose that X_1, X_2, \dots are independent and assume the values ± 1 with probability $\frac{1}{2}$ each, and let $X = \sum_{n=1}^{\infty} X_n/2^n$. Show that X is uniformly distributed over $[-1, +1]$. Calculate the characteristic functions of X and X_n and deduce (1.40). Conversely, establish (1.40) by trigonometry and conclude that X is uniformly distributed over $[-1, +1]$.

31.14. (a) Suppose that X_1, X_2, \dots are independent and assume the values 0 and 1 with probability $\frac{1}{2}$ each. Let F and G be the distribution functions of $\sum_{n=1}^{\infty} X_{2n-1}/2^{2n-1}$ and $\sum_{n=1}^{\infty} X_{2n}/2^{2n}$. Show that F and G are singular but that $F * G$ is absolutely continuous.

(b) Show that the convolution of an absolutely continuous distribution function with an arbitrary distribution function is absolutely continuous.

31.15. 31.2 \uparrow Show that the Cantor function is the distribution function of $\sum_{n=1}^{\infty} X_n/3^n$, where the X_n are independent and assume the values 0 and 2 with probability $\frac{1}{2}$ each. Express its characteristic function as an infinite product.

31.16. Show for the F of Example 31.1 that ${}_F D(1) = \infty$ and $D^F(0) = 0$ if $p_0 < \frac{1}{2}$. From (31.17) deduce that ${}_F D(x) = \infty$ and $D^F(x) = 0$ for all dyadic rationals x . Analyze the case $p_0 > \frac{1}{2}$ and sketch the graph.

31.17. 6.14 \uparrow Let F be as in Example 31.1, and let μ be the corresponding probability measure on the unit interval. Let $d_n(x)$ be the n th digit in the nonterminating binary expansion of x , and let $s_n(x) = \sum_{k=1}^n d_k(x)$. If $I_n(x)$ is the dyadic interval of order n containing x , then

$$(31.33) \quad -\frac{1}{n} \log \mu(I_n(x)) = -\left(1 - \frac{s_n(x)}{n}\right) \log p_0 - \frac{s_n(x)}{n} \log p_1.$$

(a) Show that (31.33) converges on a set of μ -measure 1 to the entropy $h = -p_0 \log p_0 - p_1 \log p_1$. From the fact that this entropy is less than $\log 2$ if $p_0 \neq \frac{1}{2}$, deduce in this case that on a set of μ -measure 1, F does not have a finite derivative.

(b) Show that (31.33) converges to $-\frac{1}{2} \log p_0 - \frac{1}{2} \log p_1$ on a set of Lebesgue measure 1. If $p_0 \neq \frac{1}{2}$ this limit exceeds $\log 2$ (arithmetic versus geometric means), and so $\mu(I_n(x))/2^{-n} \rightarrow 0$ except on a set of Lebesgue measure 0. This does not prove that $F'(x)$ exists almost everywhere, but it does show that, except for x in a set of Lebesgue measure 0, if $F'(x)$ does exist, then it is 0.

(c) Show that, if (31.33) converges to l , then

$$(31.34) \quad \lim_n \frac{\mu(I_n(x))}{(2^{-n})^\alpha} = \begin{cases} \infty & \text{if } \alpha > l/\log 2, \\ 0 & \text{if } \alpha < l/\log 2. \end{cases}$$

If (31.34) holds, then (roughly) F satisfies a Lipschitz condition[†] of (exact) order $l/\log 2$. Thus F satisfies a Lipschitz condition of order $h/\log 2$ on a set of μ -measure 1 and a Lipschitz condition of order $(-\frac{1}{2} \log p_0 - \frac{1}{2} \log p_1)/\log 2$ on a set of Lebesgue measure 1.

31.18. van der Waerden's continuous, nowhere differentiable function is $f(x) = \sum_{k=0}^{\infty} a_k(x)$, where $a_0(x)$ is the distance from x to the nearest integer and $a_k(x) = 2^{-k} a_0(2^k x)$. Show by the Weierstrass M -test that f is continuous. Use (31.8) and the ideas in Example 31.1 to show that f is nowhere differentiable.

[†]A Lipschitz condition of order α holds at x if $F(x+h) - F(x) = O(|h|^\alpha)$ as $h \rightarrow 0$; for $\alpha > 1$ this implies $F'(x) = 0$, and for $0 < \alpha < 1$ it is a smoothness condition stronger than continuity and weaker than differentiability.

- 31.19.** Show (see (31.31)) that (apart from addition of constants) a function can have only one representation $F_1 + F_2$ with F_1 absolutely continuous and F_2 singular.
- 31.20.** Show that the F_s in the Lebesgue decomposition can be further split into $F_d + F_{cs}$, where F_{cs} is continuous and singular and F_d increases only in jumps in the sense that the corresponding measure is discrete. The complete decomposition is then $F = F_{ac} + F_{cs} + F_d$.
- 31.21.** (a) Suppose that $x_1 < x_2 < \cdots$ and $\sum_n |F(x_n)| = \infty$. Show that, if F assumes the value 0 in each interval (x_n, x_{n+1}) , then it is of unbounded variation.
 (b) Define F over $[0, 1]$ by $F(0) = 0$ and $F(x) = x^\alpha \sin x^{-1}$ for $x > 0$. For which values of α is F of bounded variation?
- 31.22.** 14.4 \uparrow If f is nonnegative and Lebesgue integrable, then by Theorem 31.3 and (31.8), except for x in a set of Lebesgue measure 0,

$$(31.35) \quad \frac{1}{v-u} \int_u^v f(t) dt \rightarrow f(x)$$

if $u \leq x \leq v$, $u < v$, and $u, v \rightarrow x$. There is an analogue in which Lebesgue measure is replaced by a general probability measure μ : If f is nonnegative and integrable with respect to μ , then as $h \downarrow 0$,

$$(31.36) \quad \frac{1}{\mu(x-h, x+h]} \int_{(x-h, x+h]} f(t) \mu(dt) \rightarrow f(x)$$

on a set of μ -measure 1. Let F be the distribution function corresponding to μ , and put $\varphi(u) = \inf\{x: u \leq F(x)\}$ for $0 < u < 1$ (see (14.5)). Deduce (31.36) from (31.35) by change of variable and Problem 14.4.

SECTION 32. THE RADON-NIKODYM THEOREM

If f is a nonnegative function on a measure space $(\Omega, \mathcal{F}, \mu)$, then $\nu(A) = \int_A f d\mu$ defines another measure on \mathcal{F} . In the terminology of Section 16, ν has *density* f with respect to μ ; see (16.11). For each A in \mathcal{F} , $\mu(A) = 0$ implies that $\nu(A) = 0$. The purpose of this section is to show conversely that if this last condition holds and ν and μ are σ -finite on \mathcal{F} , then ν has a density with respect to μ . This was proved for the case $(R^1, \mathcal{R}^1, \lambda)$ in Theorems 31.7 and 31.8. The theory of the preceding section, although illuminating, is not required here.

Additive Set Functions

Throughout this section, (Ω, \mathcal{F}) is a measurable space. All sets involved are assumed as usual to lie in \mathcal{F} .

An *additive set function* is a function φ from \mathcal{F} to the reals for which

$$(32.1) \quad \varphi\left(\bigcup_n A_n\right) = \sum_n \varphi(A_n)$$

if A_1, A_2, \dots is a finite or infinite sequence of disjoint sets. A set function differs from a measure in that the values $\varphi(A)$ may be negative but must be finite—the special values $+\infty$ and $-\infty$ are prohibited. It will turn out that the series on the right in (32.1) must in fact converge absolutely, but this need not be assumed. Note that $\varphi(\emptyset) = 0$.

Example 32.1. If μ_1 and μ_2 are finite measures, then $\varphi(A) = \mu_1(A) - \mu_2(A)$ is an additive set function. It will turn out that the general additive set function has this form. A special case of this is if $\varphi(A) = \int_A f d\mu$, where f is integrable (not necessarily nonnegative). ■

The proof of the main theorem of this section (Theorem 32.2) requires certain facts about additive set functions, even though the statement of the theorem involves only measures.

Lemma 1. If $E_u \uparrow E$ or $E_u \downarrow E$, then $\varphi(E_u) \rightarrow \varphi(E)$.

PROOF. If $E_u \uparrow E$, then $\varphi(E) = \varphi(E_1 \cup \bigcup_{u=1}^{\infty} (E_{u+1} - E_u)) = \varphi(E_1) + \sum_{u=1}^{\infty} \varphi(E_{u+1} - E_u) = \lim_v [\varphi(E_1) + \sum_{u=1}^{v-1} \varphi(E_{u+1} - E_u)] = \lim_v \varphi(E_v)$ by (32.1). If $E_u \downarrow E$, then $E_u^c \uparrow E^c$, and hence $\varphi(E_u) = \varphi(\Omega) - \varphi(E_u^c) \rightarrow \varphi(\Omega) - \varphi(E^c) = \varphi(E)$. ■

Although this result is essentially the same as the corresponding ones for measures, it does require separate proof. Note that the limits need not be monotone unless φ happens to be a measure.

The Hahn Decomposition

Theorem 32.1. For any additive set function φ , there exist disjoint sets A^+ and A^- such that $A^+ \cup A^- = \Omega$, $\varphi(E) \geq 0$ for all E in A^+ , and $\varphi(E) \leq 0$ for all E in A^- .

A set A is *positive* if $\varphi(E) \geq 0$ for $E \subset A$ and *negative* if $\varphi(E) \leq 0$ for $E \subset A$. The A^+ and A^- in the theorem decompose Ω into a positive and a negative set. This is the *Hahn decomposition*.

If $\varphi(A) = \int_A f d\mu$ (see Example 32.1), the result is easy: take $A^+ = [f \geq 0]$ and $A^- = [f < 0]$.

PROOF. Let $\alpha = \sup[\varphi(A): A \in \mathcal{F}]$. Suppose that there exists a set A^+ satisfying $\varphi(A^+) = \alpha$ (which implies that α is finite). Let $A^- = \Omega - A^+$. If $A \subset A^+$ and $\varphi(A) < 0$, then $\varphi(A^+ - A) > \alpha$, an impossibility; hence A^+ is a positive set. If $A \subset A^-$ and $\varphi(A) > 0$, then $\varphi(A^+ \cup A) > \alpha$, an impossibility; hence A^- is a negative set.

It is therefore only necessary to construct a set A^+ for which $\varphi(A^+) = \alpha$. Choose sets A_n such that $\varphi(A_n) \rightarrow \alpha$, and let $A = \bigcup_n A_n$. For each n consider the 2^n sets B_{ni} (some perhaps empty) that are intersections of the form $\bigcap_{k=1}^n A'_k$, where each A'_k is either A_k or else $A - A_k$. The collection $\mathcal{B}_n = [B_{ni}: 1 \leq i \leq 2^n]$ of these sets partitions A . Clearly, \mathcal{B}_n refines \mathcal{B}_{n-1} : each B_{nj} is contained in exactly one of the $B_{n-1,i}$.

Let C_n be the union of those B_{ni} in \mathcal{B}_n for which $\varphi(B_{ni}) > 0$. Since A_n is the union of certain of the B_{ni} , it follows that $\varphi(A_n) \leq \varphi(C_n)$. Since the partitions $\mathcal{B}_1, \mathcal{B}_2, \dots$ are successively finer, $m < n$ implies that $(C_m \cup \dots \cup C_{n-1} \cup C_n) - (C_m \cup \dots \cup C_{n-1})$ is the union (perhaps empty) of certain of the sets B_{ni} ; the B_{ni} in this union must satisfy $\varphi(B_{ni}) > 0$ because they are contained in C_n . Therefore, $\varphi(C_m \cup \dots \cup C_{n-1}) \leq \varphi(C_m \cup \dots \cup C_n)$, so that by induction $\varphi(A_m) \leq \varphi(C_m) \leq \varphi(C_m \cup \dots \cup C_n)$. If $D_m = \bigcup_{n=m}^\infty C_n$, then by Lemma 1 (take $E_v = C_m \cup \dots \cup C_{m+v}$) $\varphi(A_m) \leq \varphi(D_m)$. Let $A^+ = \bigcap_{m=1}^\infty D_m$ (note that $A^+ = \limsup_n C_n$), so that $D_m \downarrow A^+$. By Lemma 1, $\alpha = \lim_m \varphi(A_m) \leq \lim_m \varphi(D_m) = \varphi(A^+)$. Thus A^+ does have maximal φ -value. ■

If $\varphi^+(A) = \varphi(A \cap A^+)$ and $\varphi^-(A) = -\varphi(A \cap A^-)$, then φ^+ and φ^- are finite measures. Thus

$$(32.2) \quad \varphi(A) = \varphi^+(A) - \varphi^-(A)$$

represents the set function φ as the difference of two finite measures having disjoint supports. If $E \subset A$, then $\varphi(E) \leq \varphi^+(E) \leq \varphi^+(A)$, and there is equality if $E = A \cap A^+$. Therefore, $\varphi^+(A) = \sup_{E \subset A} \varphi(E)$. Similarly, $\varphi^-(A) = -\inf_{E \subset A} \varphi(E)$. The measures φ^+ and φ^- are called the *upper* and *lower variations* of φ , and the measure $|\varphi|$ with value $\varphi^+(A) + \varphi^-(A)$ at A is called the *total variation*. The representation (32.2) is the *Jordan decomposition*.

Absolute Continuity and Singularity

Measures μ and ν on (Ω, \mathcal{F}) are by definition *mutually singular* if they have disjoint supports—that is, if there exist sets S_μ and S_ν such that

$$(32.3) \quad \begin{cases} \mu(\Omega - S_\mu) = 0, & \nu(\Omega - S_\nu) = 0, \\ S_\mu \cap S_\nu = \emptyset. \end{cases}$$

In this case μ is also said to be *singular with respect to* ν and ν singular with respect to μ . Note that measures are automatically singular if one of them is identically 0.

According to Theorem 31.5 a finite measure on R^1 with distribution function F is singular with respect to Lebesgue measure in the sense of (32.3) if and only if $F'(x) = 0$ except on a set of Lebesgue measure 0. In Section 31 the latter condition was taken as the definition of singularity, but of course it is the requirement of disjoint supports that can be generalized from R^1 to an arbitrary Ω .

The measure ν is *absolutely continuous* with respect to μ if for each A in \mathcal{F} , $\mu(A) = 0$ implies $\nu(A) = 0$. In this case ν is also said to be *dominated* by μ , and the relation is indicated by $\nu \ll \mu$. If $\nu \ll \mu$ and $\mu \ll \nu$, the measures are *equivalent*, indicated by $\nu \equiv \mu$.

A finite measure on the line is by Theorem 31.7 absolutely continuous in this sense with respect to Lebesgue measure if and only if the corresponding distribution function F satisfies the condition (31.28). The latter condition, taken in Section 31 as the definition of absolute continuity, is again not the one that generalizes from R^1 to Ω .

There is an ϵ - δ idea related to the definition of absolute continuity given above. Suppose that for every ϵ there exists a δ such that

$$(32.4) \quad \nu(A) < \epsilon \quad \text{if } \mu(A) < \delta.$$

If this condition holds, $\mu(A) = 0$ implies that $\nu(A) < \epsilon$ for all ϵ , and so $\nu \ll \mu$. Suppose, on the other hand, that this condition fails and that ν is finite. Then for some ϵ there exist sets A_n such that $\mu(A_n) < n^{-2}$ and $\nu(A_n) \geq \epsilon$. If $A = \limsup_n A_n$, then $\mu(A) = 0$ by the first Borel–Cantelli lemma (which applies to arbitrary measures), but $\nu(A) \geq \epsilon > 0$ by the right-hand inequality in (4.9) (which applies because ν is finite). Hence $\nu \ll \mu$ fails, and so (32.4) follows if ν is finite and $\nu \ll \mu$. If ν is finite, in order that $\nu \ll \mu$ it is therefore necessary and sufficient that for every ϵ there exist a δ satisfying (32.4). This condition is not suitable as a definition, because it need not follow from $\nu \ll \mu$ if ν is infinite.[†]

The Main Theorem

If $\nu(A) = \int_A f d\mu$, then certainly $\nu \ll \mu$. The Radon–Nikodym theorem goes in the opposite direction:

Theorem 32.2. *If μ and ν are σ -finite measures such that $\nu \ll \mu$, then there exists a nonnegative f , a density, such that $\nu(A) = \int_A f d\mu$ for all $A \in \mathcal{F}$. For two such densities f and g , $\mu[f \neq g] = 0$.*

[†]See Problem 32.3.

The uniqueness of the density up to sets of μ -measure 0 is settled by Theorem 16.10. It is only the existence that must be proved.

The density f is integrable μ if and only if ν is finite. But since f is integrable μ over A if $\nu(A) < \infty$, and since ν is assumed σ -finite, $f < \infty$ except on a set of μ -measure 0; and f can be taken finite everywhere. By Theorem 16.11, integrals with respect to ν can be calculated by the formula

$$(32.5) \quad \int_A h d\nu = \int_A hf d\mu.$$

The density whose existence is to be proved is called the *Radon-Nikodym derivative* of ν with respect to μ and is often denoted $d\nu/d\mu$. The term *derivative* is appropriate because of Theorems 31.3 and 31.8: For an absolutely continuous distribution function F on the line, the corresponding measure μ has with respect to Lebesgue measure the Radon-Nikodym derivative F' . Note that (32.5) can be written

$$(32.6) \quad \int_A h d\nu = \int_A h \frac{d\nu}{d\mu} d\mu.$$

Suppose that Theorem 32.2 holds for finite μ and ν (which is in fact enough for the probabilistic applications in the sections that follow). In the σ -finite case there is a countable decomposition of Ω into \mathcal{F} -sets A_n for which $\mu(A_n)$ and $\nu(A_n)$ are both finite. If

$$(32.7) \quad \mu_n(A) = \mu(A \cap A_n), \quad \nu_n(A) = \nu(A \cap A_n),$$

then $\nu \ll \mu$ implies $\nu_n \ll \mu_n$, and so $\nu_n(A) = \int_A f_n d\mu_n$ for some density f_n . Since μ_n has density I_{A_n} with respect to μ (Example 16.9),

$$\begin{aligned} \nu(A) &= \sum_n \nu_n(A) = \sum_n \int_A f_n d\mu_n = \sum_n \int_A f_n I_{A_n} d\mu \\ &= \int_A \sum_n f_n I_{A_n} d\mu. \end{aligned}$$

Thus $\sum_n f_n I_{A_n}$ is the density sought.

It is therefore enough to treat finite μ and ν . This requires a preliminary result.

Lemma 2. *If μ and ν are finite measures and are not mutually singular, then there exists a set A and a positive ϵ such that $\mu(A) > 0$ and $\epsilon\mu(E) \leq \nu(E)$ for all $E \subset A$.*

PROOF. Let $A_n^+ \cup A_n^-$ be a Hahn decomposition for the set function $\nu - n^{-1}\mu$; put $M = \bigcup_n A_n^+$, so that $M^c = \bigcap_n A_n^-$. Since M^c is in the negative set A_n^- for $\nu - n^{-1}\mu$, it follows that $\nu(M^c) \leq n^{-1}\mu(M^c)$; since this holds for all n , $\nu(M^c) = 0$. Thus M supports ν , and from the fact that μ and ν are not mutually singular it follows that M^c cannot support μ —that is, that $\mu(M)$ must be positive. Therefore, $\mu(A_n^+) > 0$ for some n . Take $A = A_n^+$ and $\epsilon = n^{-1}$. ■

Example 32.2. Suppose that $(\Omega, \mathcal{F}) = (R^1, \mathcal{R}^1)$, μ is Lebesgue measure λ , and $\nu(a, b] = F(b) - F(a)$. If ν and λ do not have disjoint supports, then by Theorem 31.5, $\lambda\{x: F'(x) > 0\} > 0$ and hence for some ϵ , $A = \{x: F'(x) > \epsilon\}$ satisfies $\lambda(A) > 0$. If $E = (a, b]$ is a sufficiently small interval about an x in A , then $\nu(E)/\lambda(E) = (F(b) - F(a))/(b - a) \geq \epsilon$, which is the same thing as $\epsilon\lambda(E) \leq \nu(E)$. ■

Thus Lemma 2 ties in with derivatives and quotients $\nu(E)/\mu(E)$ for “small” sets E . Martingale theory links Radon–Nikodym derivatives with such quotients; see Theorem 35.7 and Example 35.10.

PROOF OF THEOREM 32.2. Suppose that μ and ν are finite measures satisfying $\nu \ll \mu$. Let \mathcal{G} be the class of nonnegative functions g such that $\int_E g d\mu \leq \nu(E)$ for all E . If g and g' lie in \mathcal{G} , then $\max(g, g')$ also lies in \mathcal{G} because

$$\begin{aligned} \int_E \max(g, g') d\mu &= \int_{E \cap [g \geq g']} g d\mu + \int_{E \cap [g' > g]} g' d\mu \\ &\leq \nu(E \cap [g \geq g']) + \nu(E \cap [g' > g]) = \nu(E). \end{aligned}$$

Thus \mathcal{G} is closed under the formation of finite maxima. Suppose that functions g_n lie in \mathcal{G} and $g_n \uparrow g$. Then $\int_E g d\mu = \lim_n \int_E g_n d\mu \leq \nu(E)$ by the monotone convergence theorem, so that g lies in \mathcal{G} . Thus \mathcal{G} is closed under nondecreasing passages to the limit.

Let $\alpha = \sup \int g d\mu$ for g ranging over \mathcal{G} ($\alpha \leq \nu(\Omega)$). Choose g_n in \mathcal{G} so that $\int g_n d\mu > \alpha - n^{-1}$. If $f_n = \max(g_1, \dots, g_n)$ and $f = \lim f_n$, then f lies in \mathcal{G} and $\int f d\mu = \lim_n \int f_n d\mu \geq \lim_n \int g_n d\mu = \alpha$. Thus f is an element of \mathcal{G} for which $\int f d\mu$ is maximal.

Define ν_{ac} by $\nu_{ac}(E) = \int_E f d\mu$ and ν_s by $\nu_s(E) = \nu(E) - \nu_{ac}(E)$. Thus

$$(32.8) \quad \nu(E) = \nu_{ac}(E) + \nu_s(E) = \int_E f d\mu + \nu_s(E).$$

Since f is in \mathcal{G} , ν_s as well as ν_{ac} is a finite measure—that is, nonnegative. Of course, ν_{ac} is absolutely continuous with respect to μ .

Suppose that ν_s fails to be singular with respect to μ . It then follows from Lemma 2 that there are a set A and a positive ϵ such that $\mu(A) > 0$ and $\epsilon\mu(E) \leq \nu_s(E)$ for all $E \subset A$. Then for every E

$$\begin{aligned} \int_E (f + \epsilon I_A) d\mu &= \int_E f d\mu + \epsilon\mu(E \cap A) \leq \int_E f d\mu + \nu_s(E \cap A) \\ &= \int_{E \cap A} f d\mu + \nu_s(E \cap A) + \int_{E - A} f d\mu \\ &= \nu(E \cap A) + \int_{E - A} f d\mu \leq \nu(E \cap A) + \nu(E - A) \\ &= \nu(E). \end{aligned}$$

In other words, $f + \epsilon I_A$ lies in \mathcal{G} ; since $\int (f + \epsilon I_A) d\mu = \alpha + \epsilon\mu(A) > \alpha$, this contradicts the maximality of f .

Therefore, μ and ν_s are mutually singular, and there exists an S such that $\nu_s(S) = \mu(S^c) = 0$. But since $\nu \ll \mu$, $\nu_s(S^c) \leq \nu(S^c) = 0$, and so $\nu_s(\Omega) = 0$. The rightmost term in (32.8) thus drops out. ■

Absolute continuity was not used until the last step of the proof, and what the argument shows is that ν always has a decomposition (32.8) into an *absolutely continuous part* and a *singular part* with respect to μ . This is the *Lebesgue decomposition*, and it generalizes the one in the preceding section (see (31.31)).

PROBLEMS

- 32.1. There are two ways to show that the convergence in (32.1) must be absolute: Use the Jordan decomposition. Use the fact that a series converges absolutely if it has the same sum no matter what order the terms are taken in.
- 32.2. If $A^+ \cup A^-$ is a Hahn decomposition of φ , there may be other ones $A_1^+ \cup A_1^-$. Construct an example of this. Show that there is uniqueness to the extent that $\varphi(A^+ \triangle A_1^+) = \varphi(A^- \triangle A_1^-) = 0$.
- 32.3. Show that absolute continuity does not imply the ϵ - δ condition (32.4) if ν is infinite. *Hint:* Let \mathcal{F} consist of all subsets of the space of integers, let ν be counting measure, and let μ have mass n^{-2} at n . Note that μ is finite and ν is σ -finite.
- 32.4. Show that the Radon-Nikodym theorem fails if μ is not σ -finite, even if ν is finite. *Hint:* Let \mathcal{F} consist of the countable and the cocountable sets in an uncountable Ω , let μ be counting measure, and let $\nu(A)$ be 0 or 1 as A is countable or cocountable.

32.5. Let μ be the restriction of planar Lebesgue measure λ_2 to the σ -field $\mathcal{F} = \{A \times \mathbb{R}^1 : A \in \mathcal{R}^1\}$ of vertical strips. Define ν on \mathcal{F} by $\nu(A \times \mathbb{R}^1) = \lambda_2(A \times (0, 1))$. Show that ν is absolutely continuous with respect to μ but has no density. Why does this not contradict the Radon–Nikodym theorem?

32.6. Let μ , ν , and ρ be σ -finite measures on (Ω, \mathcal{F}) . Assume the Radon–Nikodym derivatives here are everywhere nonnegative and finite.

(a) Show that $\nu \ll \mu$ and $\mu \ll \rho$ imply that $\nu \ll \rho$ and

$$\frac{d\nu}{d\rho} = \frac{d\nu}{d\mu} \frac{d\mu}{d\rho}.$$

(b) Show that $\nu \equiv \mu$ implies

$$\frac{d\nu}{d\mu} = I_{[d\mu/d\nu > 0]} \left(\frac{d\mu}{d\nu} \right)^{-1}.$$

(c) Suppose that $\mu \ll \rho$ and $\nu \ll \rho$, and let A be the set where $d\nu/d\rho > 0 = d\mu/d\rho$. Show that $\nu \ll \mu$ if and only if $\rho(A) = 0$, in which case

$$\frac{d\nu}{d\mu} = I_{[d\mu/d\rho > 0]} \frac{d\nu/d\rho}{d\mu/d\rho}.$$

32.7. Show that there is a Lebesgue decomposition (32.8) in the σ -finite as well as the finite case. Prove that it is unique.

32.8. The Radon–Nikodym theorem holds if μ is σ -finite, even if ν is not. Assume at first that μ is finite (and $\nu \ll \mu$).

(a) Let \mathcal{B} be the class of (\mathcal{F} -sets) B such that $\mu(E) = 0$ or $\nu(E) = \infty$ for each $E \subset B$. Show that \mathcal{B} contains a set B_0 of maximal μ -measure.

(b) Let \mathcal{C} be the class of sets in $\Omega_0 = B_0^c$ that are countable unions of sets of finite ν -measure. Show that \mathcal{C} contains a set C_0 of maximal μ -measure. Let $D_0 = \Omega_0 - C_0$.

(c) Deduce from the maximality of B_0 and C_0 that $\mu(D_0) = \nu(D_0) = 0$.

(d) Let $\nu_0(A) = \nu(A \cap \Omega_0)$. Using the Radon–Nikodym theorem for the pair μ, ν_0 , prove it for μ, ν .

(e) Now show that the theorem holds if μ is merely σ -finite.

(f) Show that if the density can be taken everywhere finite, then ν is σ -finite.

32.9. Let μ and ν be finite measures on (Ω, \mathcal{F}) , and suppose that \mathcal{F}° is a σ -field contained in \mathcal{F} . Then the restrictions μ° and ν° of μ and ν to \mathcal{F}° are measures on $(\Omega, \mathcal{F}^\circ)$. Let $\nu_{ac}, \nu_s, \nu_{ac}^\circ, \nu_s^\circ$ be, respectively, the absolutely continuous and singular parts of ν and ν° with respect to μ and μ° . Show that $\nu_{ac}^\circ(E) \geq \nu_{ac}(E)$ and $\nu_s^\circ(E) \leq \nu_s(E)$ for $E \in \mathcal{F}^\circ$.

32.10. Suppose that μ, ν, ν_n are finite measures on (Ω, \mathcal{F}) and that $\nu(A) = \sum_n \nu_n(A)$ for all A . Let $\nu_n(A) = \int_A f_n d\mu + \nu'_n(A)$ and $\nu(A) = \int_A f d\mu + \nu'(A)$ be the decompositions (32.8); here ν' and ν'_n are singular with respect to μ . Show that $f = \sum_n f_n$ except on a set of μ -measure 0 and that $\nu'(A) = \sum_n \nu'_n(A)$ for all A . Show that $\nu \ll \mu$ if and only if $\nu_n \ll \mu$ for all n .

- 32.11.** 32.2 \uparrow Absolute continuity of a set function φ with respect to a measure μ is defined just as if φ were itself a measure: $\mu(A) = 0$ must imply that $\varphi(A) = 0$. Show that, if this holds and μ is σ -finite, then $\varphi(A) = \int_A f d\mu$ for some integrable f . Show that $A^+ = [\omega: f(\omega) \geq 0]$ and $A^- = [\omega: f(\omega) < 0]$ give a Hahn decomposition for φ . Show that the three variations satisfy $\varphi^+(A) = \int_A f^+ d\mu$, $\varphi^-(A) = \int_A f^- d\mu$, and $|\varphi|(A) = \int_A |f| d\mu$. *Hint:* To construct f , start with (32.2).
- 32.12.** \uparrow A *signed measure* φ is a set function that satisfies (32.1) if A_1, A_2, \dots are disjoint and may assume one of the values $+\infty$ and $-\infty$ but not both. Extend the Hahn and Jordan decompositions to signed measures.
- 32.13.** 31.22 \uparrow Suppose that μ and ν are a probability measure and a σ -finite measure on the line and that $\nu \ll \mu$. Show that the Radon–Nikodym derivative f satisfies

$$\lim_{h \rightarrow 0} \frac{\nu(x-h, x+h]}{\mu(x-h, x+h]} = f(x)$$

on a set of μ -measure 1.

- 32.14.** Find on the unit interval uncountably many probability measures μ_p , $0 < p < 1$, with supports S_p such that $\mu_p\{x\} = 0$ for each x and p and the S_p are disjoint in pairs.
- 32.15.** Let \mathcal{F}_0 be the field consisting of the finite and the cofinite sets in an uncountable Ω . Define φ on \mathcal{F}_0 by taking $\varphi(A)$ to be the number of points in A if A is finite, and the negative of the number of points in A^c if A is cofinite. Show that (32.1) holds (this is not true if Ω is countable). Show that there are no negative sets for φ (except the empty set), that there is no Hahn decomposition, and that φ does not have bounded range.

SECTION 33. CONDITIONAL PROBABILITY

The concepts of conditional probability and expected value with respect to a σ -field underlie much of modern probability theory. The difficulty in understanding these ideas has to do not with mathematical detail so much as with probabilistic meaning, and the way to get at this meaning is through calculations and examples, of which there are many in this section and the next.

The Discrete Case

Consider first the conditional probability of a set A with respect to another set B . It is defined of course by $P(A|B) = P(A \cap B)/P(B)$, unless $P(B)$ vanishes, in which case it is not defined at all.

It is helpful to consider conditional probability in terms of an observer in possession of partial information.[†] A probability space (Ω, \mathcal{F}, P) describes

[†]As always, *observer*, *information*, *know*, and so on are informal, nonmathematical terms; see the related discussion in Section 4 (p. 57).

the working of a mechanism, governed by chance, which produces a result ω distributed according to P ; $P(A)$ is for the observer the probability that the point ω produced lies in A . Suppose now that ω lies in B and that the observer learns this fact and no more. From the point of view of the observer, now in possession of this partial information about ω , the probability that ω also lies in A is $P(A|B)$ rather than $P(A)$. This is the idea lying back of the definition.

If, on the other hand, ω happens to lie in B^c and the observer learns of this, his probability instead becomes $P(A|B^c)$. These two conditional probabilities can be linked together by the simple function

$$(33.1) \quad f(\omega) = \begin{cases} P(A|B) & \text{if } \omega \in B, \\ P(A|B^c) & \text{if } \omega \in B^c. \end{cases}$$

The observer learns whether ω lies in B or in B^c ; his new probability for the event $\omega \in A$ is then just $f(\omega)$. Although the observer does not in general know the argument ω of f , he can calculate the value $f(\omega)$ because he knows which of B and B^c contains ω . (Note conversely that from the value $f(\omega)$ it is possible to determine whether ω lies in B or in B^c , unless $P(A|B) = P(A|B^c)$ —that is, unless A and B are independent, in which case the conditional probability coincides with the unconditional one anyway.)

The sets B and B^c partition Ω , and these ideas carry over to the general partition. Let B_1, B_2, \dots be a finite or countable partition of Ω into \mathcal{F} -sets, and let \mathcal{G} consist of all the unions of the B_i . Then \mathcal{G} is the σ -field generated by the B_i . For A in \mathcal{F} , consider the function with values

$$(33.2) \quad f(\omega) = P(A|B_i) = \frac{P(A \cap B_i)}{P(B_i)} \quad \text{if } \omega \in B_i, \quad i = 1, 2, \dots$$

If the observer learns which element B_i of the partition it is that contains ω , then his new probability for the event $\omega \in A$ is $f(\omega)$. The partition $\{B_i\}$, or equivalently the σ -field \mathcal{G} , can be regarded as an experiment, and to learn which B_i it is that contains ω is to learn the outcome of the experiment. For this reason the function or random variable f defined by (33.2) is called the *conditional probability of A given \mathcal{G}* and is denoted $P[A|\mathcal{G}]$. This is written $P[A|\mathcal{G}]_\omega$ whenever the argument ω needs to be explicitly shown.

Thus $P[A|\mathcal{G}]$ is the function whose value on B_i is the ordinary conditional probability $P(A|B_i)$. This definition needs to be completed, because $P(A|B_i)$ is not defined if $P(B_i) = 0$. In this case $P[A|\mathcal{G}]$ will be taken to have any constant value on B_i ; the value is arbitrary but must be the same over all of the set B_i . If there are nonempty sets B_i for which $P(B_i) = 0$, $P[A|\mathcal{G}]$ therefore stands for any one of a family of functions on Ω . A specific such function is for emphasis often called a *version* of the conditional

probability. Note that any two versions are equal except on a set of probability 0.

Example 33.1. Consider the Poisson process. Suppose that $0 \leq s \leq t$, and let $A = [N_s = 0]$ and $B_i = [N_t = i]$, $i = 0, 1, \dots$. Since the increments are independent (Section 23), $P(A|B_i) = P[N_s = 0]P[N_t - N_s = i]/P[N_t = i]$, and since they have Poisson distributions (see (23.9)), a simple calculation reduces this to

$$(33.3) \quad P[N_s = 0|\mathcal{G}]_\omega = \left(1 - \frac{s}{t}\right)^i \quad \text{if } \omega \in B_i, \quad i = 0, 1, 2, \dots$$

Since $i = N_t(\omega)$ on B_i , this can be written

$$(33.4) \quad P[N_s = 0|\mathcal{G}]_\omega = \left(1 - \frac{s}{t}\right)^{N_t(\omega)}.$$

Here the experiment or observation corresponding to $\{B_i\}$ or \mathcal{G} determines the number of events—telephone calls, say—occurring in the time interval $[0, t]$. For an observer who knows this number but not the locations of the calls within $[0, t]$, (33.4) gives his probability for the event that none of them occurred before time s . Although this observer does not know ω , he knows $N_t(\omega)$, which is all he needs to calculate the right side of (33.4). ■

Example 33.2. Suppose that X_0, X_1, \dots is a Markov chain with state space S as in Section 8. The events

$$(33.5) \quad [X_0 = i_0, \dots, X_n = i_n]$$

form a finite or countable partition of Ω as i_0, \dots, i_n range over S . If \mathcal{G}_n is the σ -field generated by this partition, then by the defining condition (8.2) for Markov chains, $P[X_{n+1} = j|\mathcal{G}_n]_\omega = p_{i_n j}$ holds for ω in (33.5). The sets

$$(33.6) \quad [X_n = i]$$

for $i \in S$ also partition Ω , and they generate a σ -field \mathcal{G}_n^0 smaller than \mathcal{G}_n . Now (8.2) also stipulates $P[X_{n+1} = j|\mathcal{G}_n^0]_\omega = p_{ij}$ for ω in (33.6), and the essence of the Markov property is that

$$(33.7) \quad P[X_{n+1} = j|\mathcal{G}_n] = P[X_{n+1} = j|\mathcal{G}_n^0]. \quad \blacksquare$$

The General Case

If \mathcal{G} is the σ -field generated by a partition B_1, B_2, \dots , then the general element of \mathcal{G} is a disjoint union $B_{i_1} \cup B_{i_2} \cup \dots$, finite or countable, of certain of the B_i . To know which set B_i it is that contains ω is the same thing

as to know which sets in \mathcal{G} contain ω and which do not. This second way of looking at the matter carries over to the general σ -field \mathcal{G} contained in \mathcal{F} . (As always, the probability space is (Ω, \mathcal{F}, P) .) The σ -field \mathcal{G} will not in general come from a partition as above.

One can imagine an observer who knows for each G in \mathcal{G} whether $\omega \in G$ or $\omega \in G^c$. Thus the σ -field \mathcal{G} can in principle be identified with an experiment or observation. This is the point of view adopted in Section 4; see p. 57. It is natural to try and define conditional probabilities $P[A|\mathcal{G}]$ with respect to the experiment \mathcal{G} . To do this, fix an A in \mathcal{F} and define a finite measure ν on \mathcal{G} by

$$\nu(G) = P(A \cap G), \quad G \in \mathcal{G}.$$

Then $P(G) = 0$ implies that $\nu(G) = 0$. The Radon–Nikodym theorem can be applied to the measures ν and P on the measurable space (Ω, \mathcal{G}) because the first one is absolutely continuous with respect to the second.[†] It follows that there exists a function or random variable f , measurable \mathcal{G} and integrable with respect to P , such that[†] $P(A \cap G) = \nu(G) = \int_G f dP$ for all G in \mathcal{G} .

Denote this function f by $P[A|\mathcal{G}]$. It is a random variable with two properties:

- (i) $P[A|\mathcal{G}]$ is measurable \mathcal{G} and integrable.
- (ii) $P[A|\mathcal{G}]$ satisfies the functional equation

$$(33.8) \quad \int_G P[A|\mathcal{G}] dP = P(A \cap G), \quad G \in \mathcal{G}.$$

There will in general be many such random variables $P[A|\mathcal{G}]$, but any two of them are equal with probability 1. A specific such random variable is called a *version* of the conditional probability.

If \mathcal{G} is generated by a partition B_1, B_2, \dots the function f defined by (33.2) is measurable \mathcal{G} because $[\omega: f(\omega) \in H]$ is the union of those B_i over which the constant value of f lies in H . Any G in \mathcal{G} is a disjoint union $G = \bigcup_k B_{i_k}$, and $P(A \cap G) = \sum_k P(A|B_{i_k})P(B_{i_k})$, so that (33.2) satisfies (33.8) as well. Thus the general definition is an extension of the one for the discrete case.

Condition (i) in the definition above in effect requires that the values of $P[A|\mathcal{G}]$ depend only on the sets in \mathcal{G} . An observer who knows the outcome of \mathcal{G} viewed as an experiment knows for each G in \mathcal{G} whether it contains ω or not; for each x he knows this in particular for the set $[\omega': P[A|\mathcal{G}]_{\omega'} = x]$,

[†]Let P_0 be the restriction of P to \mathcal{G} (Example 10.4), and find on (Ω, \mathcal{G}) a density f for ν with respect to P_0 . Then, for $G \in \mathcal{G}$, $\nu(G) = \int_G f dP_0 = \int_G f dP$ (Example 16.4). If g is another such density, then $P[f \neq g] = P_0[f \neq g] = 0$.

and hence he knows in principle the functional value $P[A|\mathcal{G}]_\omega$ even if he does not know ω itself. In Example 33.1 a knowledge of $N_i(\omega)$ suffices to determine the value of (33.4)— ω itself is not needed.

Condition (ii) in the definition has a gambling interpretation. Suppose that the observer, after he has learned the outcome of \mathcal{G} , is offered the opportunity to bet on the event A (unless A lies in \mathcal{G} , he does not yet know whether or not it occurred). He is required to pay an entry fee of $P[A|\mathcal{G}]$ units and will win 1 unit if A occurs and nothing otherwise. If the observer decides to bet and pays his fee, he gains $1 - P[A|\mathcal{G}]$ if A occurs and $-P[A|\mathcal{G}]$ otherwise, so that his gain is

$$(1 - P[A|\mathcal{G}])I_A + (-P[A|\mathcal{G}])I_{A^c} = I_A - P[A|\mathcal{G}].$$

If he declines to bet, his gain is of course 0. Suppose that he adopts the strategy of betting if G occurs but not otherwise, where G is some set in \mathcal{G} . He can actually carry out this strategy, since after learning the outcome of the experiment \mathcal{G} he knows whether or not G occurred. His expected gain with this strategy is his gain integrated over G :

$$\int_G (I_A - P[A|\mathcal{G}]) dP.$$

But (33.8) is exactly the requirement that this vanish for each G in \mathcal{G} . Condition (ii) requires then that each strategy be fair in the sense that the observer stands neither to win nor to lose on the average. Thus $P[A|\mathcal{G}]$ is the just entry fee, as intuition requires.

Example 33.3. Suppose that $A \in \mathcal{G}$, which will always hold if \mathcal{G} coincides with the whole σ -field \mathcal{F} . Then I_A satisfies conditions (i) and (ii), so that $P[A|\mathcal{G}] = I_A$ with probability 1. If $A \in \mathcal{G}$, then to know the outcome of \mathcal{G} viewed as an experiment is in particular to know whether or not A has occurred. ■

Example 33.4. If \mathcal{G} is $\{0, \Omega\}$, the smallest possible σ -field, every function measurable \mathcal{G} must be constant. Therefore, $P[A|\mathcal{G}]_\omega = P(A)$ for all ω in this case. The observer learns nothing from the experiment \mathcal{G} . ■

According to these two examples, $P[A|\{0, \Omega\}]$ is identically $P(A)$, whereas I_A is a version of $P[A|\mathcal{F}]$. For any \mathcal{G} , the function identically equal to $P(A)$ satisfies condition (i) in the definition of conditional probability, whereas I_A satisfies condition (ii). Condition (i) becomes more stringent as \mathcal{G} decreases, and condition (ii) becomes more stringent as \mathcal{G} increases. The two conditions work in opposite directions and between them delimit the class of versions of $P[A|\mathcal{G}]$.

Example 33.5. Let Ω be the plane R^2 and let \mathcal{F} be the class \mathcal{R}^2 of planar Borel sets. A point of Ω is a pair (x, y) of reals. Let \mathcal{G} be the σ -field consisting of the vertical strips, the product sets $E \times R^1 = [(x, y): x \in E]$, where E is a linear Borel set. If the observer knows for each strip $E \times R^1$ whether or not it contains (x, y) , then, as he knows this for each one-point set E , he knows the value of x . Thus the experiment \mathcal{G} consists in the determination of the first coordinate of the sample point. Suppose now that P is a probability measure on \mathcal{R}^2 having a density $f(x, y)$ with respect to planar Lebesgue measure: $P(A) = \iint_A f(x, y) dx dy$. Let A be a horizontal strip $R^1 \times F = [(x, y): y \in F]$, F being a linear Borel set. The conditional probability $P[A|\mathcal{G}]$ can be calculated explicitly.

Put

$$(33.9) \quad \varphi(x, y) = \frac{\int_F f(x, t) dt}{\int_{R^1} f(x, t) dt}.$$

Set $\varphi(x, y) = 0$, say, at points where the denominator here vanishes; these points form a set of P -measure 0. Since $\varphi(x, y)$ is a function of x alone, it is measurable \mathcal{G} . The general element of \mathcal{G} being $E \times R^1$, it will follow that φ is a version of $P[A|\mathcal{G}]$ if it is shown that

$$(33.10) \quad \int_{E \times R^1} \varphi(x, y) dP(x, y) = P(A \cap (E \times R^1)).$$

Since $A = R^1 \times F$, the right side here is $P(E \times F)$. Since P has density f , Theorem 16.11 and Fubini's theorem reduce the left side to

$$\begin{aligned} \int_E \left\{ \int_{R^1} \varphi(x, y) f(x, y) dy \right\} dx &= \int_E \left\{ \int_F f(x, t) dt \right\} dx \\ &= \iint_{E \times F} f(x, y) dx dy = P(E \times F). \end{aligned}$$

Thus (33.9) does give a version of $P[R^1 \times F|\mathcal{G}]$. ■

The right side of (33.9) is the classical formula for the conditional probability of the event $R^1 \times F$ (the event that $y \in F$) given the event $\{x\} \times R^1$ (given the value of x). Since the event $\{x\} \times R^1$ has probability 0, the formula $P(A|B) = P(A \cap B)/P(B)$ does not work here. The whole point of this section is the systematic development of a notion of conditional probability that covers conditioning with respect to events of probability 0. This is accomplished by conditioning with respect to *collections* of events—that is, with respect to σ -fields \mathcal{G} .

Example 33.6. The set A is by definition independent of the σ -field \mathcal{G} if it is independent of each G in \mathcal{G} : $P(A \cap G) = P(A)P(G)$. This being the same thing as $P(A \cap G) = \int_G P(A) dP$, A is independent of \mathcal{G} if and only if $P[A|\mathcal{G}] = P(A)$ with probability 1. ■

The σ -field $\sigma(X)$ generated by a random variable X consists of the sets $[\omega: X(\omega) \in H]$ for $H \in \mathcal{R}^1$; see Theorem 20.1. The conditional probability of A given X is defined as $P[A|\sigma(X)]$ and is denoted $P[A|X]$. Thus $P[A|X] = P[A|\sigma(X)]$ by definition. From the experiment corresponding to the σ -field $\sigma(X)$, one learns which of the sets $[\omega': X(\omega') = x]$ contains ω and hence learns the value $X(\omega)$. Example 33.5 is a case of this: take $X(x, y) = x$ for (x, y) in the sample space $\Omega = R^2$ there.

This definition applies without change to random vector, or, equivalently, to a finite set of random variables. It can be adapted to arbitrary sets of random variables as well. For any such set $[X_t, t \in T]$, the σ -field $\sigma[X_t, t \in T]$ it generates is the smallest σ -field with respect to which each X_t is measurable. It is generated by the collection of sets of the form $[\omega: X_t(\omega) \in H]$ for t in T and H in \mathcal{R}^1 . The conditional probability $P[A|X_t, t \in T]$ of A with respect to this set of random variables is by definition the conditional probability $P[A|\sigma[X_t, t \in T]]$ of A with respect to the σ -field $\sigma[X_t, t \in T]$.

In this notation the property (33.7) of Markov chains becomes

$$(33.11) \quad P[X_{n+1} = j | X_0, \dots, X_n] = P[X_{n+1} = j | X_n].$$

The conditional probability of $[X_{n+1} = j]$ is the same for someone who knows the present state X_n as for someone who knows the present state X_n and the past states X_0, \dots, X_{n-1} as well.

Example 33.7. Let X and Y be random vectors of dimensions j and k , let μ be the distribution of X over R^j , and suppose that X and Y are independent. According to (20.30),

$$P[X \in H, (X, Y) \in J] = \int_H P[(x, Y) \in J] \mu(dx)$$

for $H \in \mathcal{R}^j$ and $J \in \mathcal{R}^{j+k}$. This is a consequence of Fubini's theorem; it has a conditional-probability interpretation. For each x in R^j put

$$(33.12) \quad f(x) = P[(x, Y) \in J] = P[\omega': (x, Y(\omega')) \in J].$$

By Theorem 20.1(ii), $f(X(\omega))$ is measurable $\sigma(X)$, and since μ is the distribution of X , a change of variable gives

$$\int_{[X \in H]} f(X(\omega)) P(d\omega) = \int_H f(x) \mu(dx) = P([(X, Y) \in J] \cap [X \in H]).$$

Since $[X \in H]$ is the general element of $\sigma(X)$, this proves that

$$(33.13) \quad f(X(\omega)) = P[(X, Y) \in J | X]_{\omega}$$

with probability 1. ■

The fact just proved can be written

$$\begin{aligned} P[(X, Y) \in J | X]_{\omega} &= P[(X(\omega), Y) \in J] \\ &= P[\omega': (X(\omega), Y(\omega')) \in J]. \end{aligned}$$

Replacing ω' by ω on the right here causes a notational collision like the one replacing y by x causes in $\int_a^b f(x, y) dy$.

Suppose that X and Y are independent random variables and that Y has distribution function F . For $J = [(u, v): \max\{u, v\} \leq m]$, (33.12) is 0 for $m < x$ and $F(m)$ for $m \geq x$; if $M = \max\{X, Y\}$, then (33.13) gives

$$(33.14) \quad P[M \leq m | X]_{\omega} = I_{[X \leq m]}(\omega) F(m)$$

with probability 1. All equations involving conditional probabilities must be qualified in this way by the phrase *with probability 1*, because the conditional probability is unique only to within a set of probability 0.

The following theorem is useful for checking conditional probabilities.

Theorem 33.1. *Let \mathcal{P} be a π -system generating the σ -field \mathcal{G} , and suppose that Ω is a finite or countable union of sets in \mathcal{P} . An integrable function f is a version of $P[A | \mathcal{G}]$ if it is measurable \mathcal{G} and if*

$$(33.15) \quad \int_G f dP = P(A \cap G)$$

holds for all G in \mathcal{P} .

PROOF. Apply Theorem 10.4. ■

The condition that Ω is a finite or countable union of \mathcal{P} sets cannot be suppressed; see Example 10.5.

Example 33.8. Suppose that X and Y are independent random variables with a common distribution function F that is positive and continuous. What is the conditional probability of $[X \leq x]$ given the random variable $M = \max\{X, Y\}$? As it should clearly be 1 if $M \leq x$, suppose that $M > x$. Since $X \leq x$ requires $M = Y$, the chance of which is $\frac{1}{2}$ by symmetry, the conditional probability of $[X \leq x]$ should by independence be $\frac{1}{2}F(x)/F(m) = \frac{1}{2}P[X \leq x | X \leq m]$ with the random variable M substituted

for m . Intuition thus gives

$$(33.16) \quad P[X \leq x \| M]_{\omega} = I_{\{M \leq x\}}(\omega) + \frac{1}{2} I_{\{M > x\}}(\omega) \frac{F(x)}{F(M(\omega))}.$$

It suffices to check (33.15) for sets $G = [M \leq m]$, because these form a π -system generating $\sigma(M)$. The functional equation reduces to

$$(33.17) \quad P[M \leq \min\{x, m\}] + \frac{1}{2} \int_{x < M \leq m} \frac{F(x)}{F(M)} dP = P[M \leq m, X \leq x].$$

Since the other case is easy, suppose that $x < m$. Since the distribution of (X, Y) is product measure, it follows by Fubini's theorem and the assumed continuity of F that

$$\begin{aligned} \int_{x < M \leq m} \frac{1}{F(M)} dP &= \iint_{\substack{u \leq v \\ x < v \leq m}} \frac{1}{F(v)} dF(u) dF(v) \\ &\quad + \iint_{\substack{v < u \\ x < u \leq m}} \frac{1}{F(u)} dF(u) dF(v) = 2(F(m) - F(x)), \end{aligned}$$

which gives (33.17). ■

Example 33.9. A collection $[X_t: t \geq 0]$ of random variables is a *Markov process in continuous time* if for $k \geq 1$, $0 \leq t_1 \leq \cdots \leq t_k \leq u$, and $H \in \mathcal{R}^1$,

$$(33.18) \quad P[X_u \in H \| X_{t_1}, \dots, X_{t_k}] = P[X_u \in H \| X_{t_k}]$$

holds with probability 1. The analogue for discrete time is (33.11). (The X_n there have countable range as well, and the transition probabilities are constant in time, conditions that are not imposed here.)

Suppose that $t \leq u$. Looking on the right side of (33.18) as a version of the conditional probability on the left shows that

$$(33.19) \quad \int_G P[X_u \in H \| X_t] dP = P([X_u \in H] \cap G)$$

if $0 \leq t_1 \leq \cdots \leq t_k = t \leq u$ and $G \in \sigma(X_{t_1}, \dots, X_{t_k})$. Fix t , u , and H , and let k and t_1, \dots, t_k vary. Consider the class $\mathcal{P} = \bigcup \sigma(X_{t_1}, \dots, X_{t_k})$, the union extending over all $k \geq 1$ and all k -tuples satisfying $0 \leq t_1 \leq \cdots \leq t_k = t$. If $A \in \sigma(X_{t_1}, \dots, X_{t_k})$ and $B \in \sigma(X_{s_1}, \dots, X_{s_j})$, then $A \cap B \in \sigma(X_{r_1}, \dots, X_{r_j})$, where the r_α are the s_β and the t_γ merged together. Thus \mathcal{P} is a π -system. Since \mathcal{P} generates $\sigma[X_s: s \leq t]$ and $P[X_u \in H \| X_t]$ is measurable with respect to this σ -field, it follows by (33.19) and Theorem 33.1 that $P[X_u \in H \| X_t]$ is a version of $P[X_u \in H \| X_s, s \leq t]$:

$$(33.20) \quad P[X_u \in H \| X_s, s \leq t] = P[X_u \in H \| X_t], \quad t \leq u,$$

with probability 1.

This says that for calculating conditional probabilities about the future, the present $\sigma(X_t)$ is equivalent to the present and the *entire* past $\sigma[X_s: s \leq t]$. This follows from the apparently weaker condition (33.18). ■

Example 33.10. The Poisson process $[N_t: t \geq 0]$ has independent increments (Section 23). Suppose that $0 \leq t_1 \leq \cdots \leq t_k \leq u$. The random vector $(N_{t_1}, N_{t_2} - N_{t_1}, \dots, N_{t_k} - N_{t_{k-1}})$ is independent of $N_u - N_{t_k}$, and so (Theorem 20.2) $(N_{t_1}, N_{t_2}, \dots, N_{t_k})$ is independent of $N_u - N_{t_k}$. If J is the set of points (x_1, \dots, x_k, y) in R^{k+1} such that $x_k + y \in H$, where $H \in \mathcal{R}^1$, and if ν is the distribution of $N_u - N_{t_k}$, then (33.12) is $P[(x_1, \dots, x_k, N_u - N_{t_k}) \in J] = P[x_k + N_u - N_{t_k} \in H] = \nu(H - x_k)$. Therefore, (33.13) gives $P[N_u \in H | N_{t_1}, \dots, N_{t_k}] = \nu(H - N_{t_k})$. This holds also if $k = 1$, and hence $P[N_u \in H | N_{t_1}, \dots, N_{t_k}] = P[N_u \in H | N_{t_k}]$. The Poisson process thus has the Markov property (33.18); this is a consequence solely of the independence of the increments. The extended Markov property (33.20) follows. ■

Properties of Conditional Probability

Theorem 33.2. With probability 1, $P[\emptyset | \mathcal{G}] = 0$, $P[\Omega | \mathcal{G}] = 1$; and

$$(33.21) \quad 0 \leq P[A | \mathcal{G}] \leq 1$$

for each A . If A_1, A_2, \dots is a finite or countable sequence of disjoint sets, then

$$(33.22) \quad P\left[\bigcup_n A_n | \mathcal{G}\right] = \sum_n P[A_n | \mathcal{G}]$$

with probability 1.

PROOF. For each version of the conditional probability, $\int_G P[A | \mathcal{G}] dP = P(A \cap G) \geq 0$ for each G in \mathcal{G} ; since $P[A | \mathcal{G}]$ is measurable \mathcal{G} , it must be nonnegative except on a set of P -measure 0. The other inequality in (33.21) is proved the same way.

If the A_n are disjoint and if G lies in \mathcal{G} , it follows (Theorem 16.6) that

$$\begin{aligned} \int_G \left(\sum_n P[A_n | \mathcal{G}] \right) dP &= \sum_n \int_G P[A_n | \mathcal{G}] dP = \sum_n P(A_n \cap G) \\ &= P\left(\left(\bigcup_n A_n\right) \cap G\right). \end{aligned}$$

Thus $\sum_n P[A_n | \mathcal{G}]$, which is certainly measurable \mathcal{G} , satisfies the functional equation for $P[\bigcup_n A_n | \mathcal{G}]$, and so must coincide with it except perhaps on a set of P -measure 0. Hence (33.22). ■

Additional useful facts can be established by similar arguments. If $A \subset B$, then

$$(33.23) \quad P[B - A | \mathcal{G}] = P[B | \mathcal{G}] - P[A | \mathcal{G}], \quad P[A | \mathcal{G}] \leq P[B | \mathcal{G}].$$

The inclusion-exclusion formula

$$(33.24) \quad P\left[\bigcup_{i=1}^n A_i | \mathcal{G}\right] = \sum_i P[A_i | \mathcal{G}] - \sum_{i < j} P[A_i \cap A_j | \mathcal{G}] + \cdots$$

holds. If $A_n \uparrow A$, then

$$(33.25) \quad P[A_n | \mathcal{G}] \uparrow P[A | \mathcal{G}],$$

and if $A_n \downarrow A$, then

$$(33.26) \quad P[A_n | \mathcal{G}] \downarrow P[A | \mathcal{G}].$$

Further, $P(A) = 1$ implies that

$$(33.27) \quad P[A | \mathcal{G}] = 1,$$

and $P(A) = 0$ implies that

$$(33.28) \quad P[A | \mathcal{G}] = 0.$$

Of course (33.23) through (33.28) hold with probability 1 only.

Difficulties and Curiosities

This section has been devoted almost entirely to examples connecting the abstract definition (33.8) with the probabilistic idea lying back of it. There are pathological examples showing that the interpretation of conditional probability in terms of an observer with partial information breaks down in certain cases.

Example 33.11. Let (Ω, \mathcal{F}, P) be the unit interval Ω with Lebesgue measure P on the σ -field \mathcal{F} of Borel subsets of Ω . Take \mathcal{G} to be the σ -field of sets that are either countable or cocountable. Then the function identically equal to $P(A)$ is a version of $P[A | \mathcal{G}]$: (33.8) holds because $P(G)$ is either 0 or 1 for every G in \mathcal{G} . Therefore,

$$(33.29) \quad P[A | \mathcal{G}]_\omega = P(A)$$

with probability 1. But since \mathcal{G} contains all one-point sets, to know which

elements of \mathcal{G} contain ω is to know ω itself. Thus \mathcal{G} viewed as an experiment should be completely informative—the observer given the information in \mathcal{G} should know ω exactly—and so one might expect that

$$(33.30) \quad P[A|\mathcal{G}]_{\omega} = \begin{cases} 1 & \text{if } \omega \in A, \\ 0 & \text{if } \omega \notin A. \end{cases}$$

This is Example 4.10 in a new form. ■

The mathematical definition gives (33.29); the heuristic considerations lead to (33.30). Of course, (33.29) is right and (33.30) is wrong. The heuristic view breaks down in certain cases but is nonetheless illuminating and cannot, since it does not intervene in proofs, lead to any difficulties.

The point of view in this section has been “global.” To each fixed A in \mathcal{F} has been attached a function (usually a family of functions) $P[A|\mathcal{G}]_{\omega}$ defined over all of Ω . What happens if the point of view is reversed—if ω is fixed and A varies over \mathcal{F} ? Will this result in a probability measure on \mathcal{F} ? Intuition says it should, and if it does, then (33.21) through (33.28) all reduce to standard facts about measures.

Suppose that B_1, \dots, B_r is a partition of Ω into \mathcal{F} -sets, and let $\mathcal{G} = \sigma(B_1, \dots, B_r)$. If $P(B_1) = 0$ and $P(B_i) > 0$ for the other i , then one version of $P[A|\mathcal{G}]$ is

$$P[A|\mathcal{G}]_{\omega} = \begin{cases} 17 & \text{if } \omega \in B_1, \\ \frac{P(A \cap B_i)}{P(B_i)} & \text{if } \omega \in B_i, i = 2, \dots, r. \end{cases}$$

With this choice of version for each A , $P[A|\mathcal{G}]_{\omega}$ is, as a function of A , a probability measure on \mathcal{F} if $\omega \in B_2 \cup \dots \cup B_r$, but not if $\omega \in B_1$. The “wrong” versions have been chosen. If, for example,

$$P[A|\mathcal{G}]_{\omega} = \begin{cases} P(A) & \text{if } \omega \in B_1, \\ \frac{P(A \cap B_i)}{P(B_i)} & \text{if } \omega \in B_i, i = 2, \dots, r, \end{cases}$$

then $P[A|\mathcal{G}]_{\omega}$ is a probability measure in A for each ω . Clearly, versions such as this one exist if \mathcal{G} is finite.

It might be thought that for an arbitrary σ -field \mathcal{G} in \mathcal{F} versions of the various $P[A|\mathcal{G}]$ can be so chosen that $P[A|\mathcal{G}]_{\omega}$ is for each fixed ω a probability measure as A varies over \mathcal{F} . It is possible to construct a

counterexample showing that this is not so.[†] The example is possible because the exceptional ω -set of probability 0 where (33.22) fails depends on the sequence A_1, A_2, \dots ; if there are uncountably many such sequences, it can happen that the union of these exceptional sets has positive probability whatever versions $P[A|\mathcal{G}]$ are chosen.

The existence of such pathological examples turns out not to matter. Example 33.9 illustrates the reason why. From the assumption (33.18) the notably stronger conclusion (33.20) was reached. Since the set $[X_u \in H]$ is fixed throughout the argument, it does not matter that conditional probabilities may not, in fact, be measures. What does matter for the theory is Theorem 33.2 and its extensions.

Consider a point ω_0 with the property that $P(G) > 0$ for every G in \mathcal{G} that contains ω_0 . This will be true if the one-point set $\{\omega_0\}$ lies in \mathcal{F} and has positive probability. Fix any versions of the $P[A|\mathcal{G}]$. For each A the set $[\omega: P[A|\mathcal{G}]_\omega < 0]$ lies in \mathcal{G} and has probability 0; it therefore cannot contain ω_0 . Thus $P[A|\mathcal{G}]_{\omega_0} \geq 0$. Similarly, $P[\Omega|\mathcal{G}]_{\omega_0} = 1$, and, if the A_n are disjoint, $P[\bigcup_n A_n|\mathcal{G}]_{\omega_0} = \sum_n P[A_n|\mathcal{G}]_{\omega_0}$. Therefore, $P[A|\mathcal{G}]_{\omega_0}$ is a probability measure as A ranges over \mathcal{F} .

Thus conditional probabilities behave like probabilities at points of positive probability. That they may not do so at points of probability 0 causes no problem because individual such points have no effect on the probabilities of sets. Of course, *sets* of points individually having probability 0 do have an effect, but here the global point of view reenters.

Conditional Probability Distributions

Let X be a random variable on (Ω, \mathcal{F}, P) , and let \mathcal{G} be a σ -field in \mathcal{F} .

Theorem 33.3. *There exists a function $\mu(H, \omega)$, defined for H in \mathcal{R}^1 and ω in Ω , with these two properties:*

- (i) *For each ω in Ω , $\mu(\cdot, \omega)$ is a probability measure on \mathcal{R}^1 .*
- (ii) *For each H in \mathcal{R}^1 , $\mu(H, \cdot)$ is a version of $P[X \in H|\mathcal{G}]$.*

The probability measure $\mu(\cdot, \omega)$ is a *conditional distribution* of X given \mathcal{G} . If $\mathcal{G} = \sigma(Z)$, it is a conditional distribution of X given Z .

PROOF. For each rational r , let $F(r, \omega)$ be a version of $P[X \leq r|\mathcal{G}]_\omega$. If $r \leq s$, then by (33.23),

$$(33.31) \quad F(r, \omega) \leq F(s, \omega)$$

[†]The argument is outlined in Problem 33.11. It depends on the construction of certain nonmeasurable sets.

for ω outside a \mathcal{G} -set A_{rs} of probability 0. By (33.26),

$$(33.32) \quad F(r, \omega) = \lim_n F(r + n^{-1}, \omega)$$

for ω outside a \mathcal{G} -set B_r of probability 0. Finally, by (33.25) and (33.26),

$$(33.33) \quad \lim_{r \rightarrow -\infty} F(r, \omega) = 0, \quad \lim_{r \rightarrow \infty} F(r, \omega) = 1$$

outside a \mathcal{G} -set C of probability 0. As there are only countably many of these exceptional sets, their union E lies in \mathcal{G} and has probability 0.

For $\omega \notin E$ extend $F(\cdot, \omega)$ to all of R^1 by setting $F(x, \omega) = \inf\{F(r, \omega) : x < r\}$. For $\omega \in E$ take $F(x, \omega) = F(x)$, where F is some arbitrary but fixed distribution function. Suppose that $\omega \notin E$. By (33.31) and (33.32), $F(x, \omega)$ agrees with the first definition on the rationals and is nondecreasing; it is right-continuous; and by (33.33) it is a probability distribution function. Therefore, there exists a probability measure $\mu(\cdot, \omega)$ on (R^1, \mathcal{R}^1) with distribution function $F(\cdot, \omega)$. For $\omega \in E$, let $\mu(\cdot, \omega)$ be the probability measure corresponding to $F(x)$. Then condition (i) is satisfied.

The class of H for which $\mu(H, \cdot)$ is measurable \mathcal{G} is a λ -system containing the sets $H = (-\infty, r]$ for rational r ; therefore $\mu(H, \cdot)$ is measurable \mathcal{G} for H in \mathcal{R}^1 .

By construction, $\mu((-\infty, r], \omega) = P[X \leq r | \mathcal{G}]_\omega$ with probability 1 for rational r ; that is, for $H = (-\infty, r]$ as well as for $H = R^1$,

$$\int_G \mu(H, \omega) P(d\omega) = P([X \in H] \cap G)$$

for all G in \mathcal{G} . Fix G . Each side of this equation is a measure as a function of H , and so the equation must hold for all H in \mathcal{R}^1 . ■

Example 33.12. Let X and Y be random variables whose joint distribution ν in R^2 has density $f(x, y)$ with respect to Lebesgue measure: $P[(X, Y) \in A] = \nu(A) = \iint_A f(x, y) dx dy$. Let $g(x, y) = f(x, y) / \int_{R^1} f(x, t) dt$, and let $\mu(H, x) = \int_H g(x, y) dy$ have probability density $g(x, \cdot)$; if $\int_{R^1} f(x, t) dt = 0$, let $\mu(\cdot, x)$ be an arbitrary probability measure on the line. Then $\mu(H, X(\omega))$ will serve as the conditional distribution of Y given X . Indeed, (33.10) is the same thing as $\int_{E \times R^1} \mu(F, x) d\nu(x, y) = \nu(E \times F)$, and a change of variable gives $\int_{[X \in E]} \mu(F, X(\omega)) P(d\omega) = P[X \in E, Y \in F]$. Thus $\mu(F, X(\omega))$ is a version of $P[Y \in F | X]_\omega$. This is a new version of Example 33.5. ■

PROBLEMS

33.1. 20.27↑ *Borel's paradox.* Suppose that a random point on the sphere is specified by longitude Θ and latitude Φ , but restrict Θ by $0 \leq \Theta < \pi$, so that Θ specifies the complete meridian circle (not semicircle) containing the point, and compensate by letting Φ range over $(-\pi, \pi]$.

(a) Show that for given Θ the conditional distribution of Φ has density $\frac{1}{4}|\cos \phi|$ over $(-\pi, +\pi]$. If the point lies on, say, the meridian circle through Greenwich, it is therefore not uniformly distributed over that great circle.

(b) Show that for given Φ the conditional distribution of Θ is uniform over $(0, \pi)$. If the point lies on the equator (Φ is 0 or π), it is therefore uniformly distributed over that great circle.

Since the point is uniformly distributed over the spherical surface and great circles are indistinguishable, (a) and (b) stand in apparent contradiction. This shows again the inadmissibility of conditioning with respect to an isolated event of probability 0. The relevant σ -field must not be lost sight of.

33.2. 20.16↑ Let X and Y be independent, each having the standard normal distribution, and let (R, Θ) be the polar coordinates for (X, Y) .

(a) Show that $X + Y$ and $X - Y$ are independent and that $R^2 = [(X + Y)^2 + (X - Y)^2]/2$, and conclude that the conditional distribution of R^2 given $X - Y$ is the chi-squared distribution with one degree of freedom translated by $(X - Y)^2/2$.

(b) Show that the conditional distribution of R^2 given Θ is chi-squared with two degrees of freedom.

(c) If $X - Y = 0$, the conditional distribution of R^2 is chi-squared with one degree of freedom. If $\Theta = \pi/4$ or $\Theta = 5\pi/4$, the conditional distribution of R^2 is chi-squared with two degrees of freedom. But the events $[X - Y = 0]$ and $[\Theta = \pi/4] \cup [\Theta = 5\pi/4]$ are the same. Resolve the apparent contradiction.

33.3. ↑ Paradoxes of a somewhat similar kind arise in very simple cases.

(a) Of three prisoners, call them 1, 2, and 3, two have been chosen by lot for execution. Prisoner 3 says to the guard, "Which of 1 and 2 is to be executed? One of them will be, and you give me no information about myself in telling me which it is." The guard finds this reasonable and says, "Prisoner 1 is to be executed." And now 3 reasons, "I know that 1 is to be executed; the other will be either 2 or me, and so my chance of being executed is now only $\frac{1}{2}$, instead of the $\frac{2}{3}$ it was before," Apparently, the guard *has* given him information.

If one looks for a σ -field, it must be the one describing the guard's answer, and it then becomes clear that the sample space is incompletely specified. Suppose that, if 1 and 2 are to be executed, the guard's response is "1" with probability p and "2" with probability $1 - p$; and, of course, suppose that, if 3 is to be executed, the guard names the other victim. Calculate the conditional probabilities.

(b) Assume that among families with two children the four sex distributions are equally likely. You have been introduced to one of the two children in such a family, and he is a boy. What is the conditional probability that the other is a boy as well?

- 33.4. (a) Consider probability spaces (Ω, \mathcal{F}, P) and $(\Omega', \mathcal{F}', P')$; suppose that $T: \Omega \rightarrow \Omega'$ is measurable \mathcal{F}/\mathcal{F}' and $P' = PT^{-1}$. Let \mathcal{G}' be a σ -field in \mathcal{F}' , and take \mathcal{G} to be the σ -field $[T^{-1}G': G' \in \mathcal{G}']$. For $A' \in \mathcal{F}'$, show by (16.18) that $P[T^{-1}A'|\mathcal{G}]_\omega = P'[A'|\mathcal{G}']_{T\omega}$ with P -probability 1.
- (b) Now take $(\Omega', \mathcal{F}', P') = (R^2, \mathcal{B}^2, \mu)$, where μ is the distribution of a random vector (X, Y) on (Ω, \mathcal{F}, P) . Suppose that (X, Y) has density f , and show by (33.9) that

$$P[Y \in F | X]_\omega = \frac{\int_F f(X(\omega), t) dt}{\int_{R^1} f(X(\omega), t) dt}$$

with probability 1.

- 33.5. \uparrow (a) There is a slightly different approach to conditional probability. Let (Ω, \mathcal{F}, P) be a probability space, (Ω', \mathcal{F}') a measurable space, and $T: \Omega \rightarrow \Omega'$ a mapping measurable \mathcal{F}/\mathcal{F}' . Define a measure ν on \mathcal{F}' by $\nu(A') = P(A \cap T^{-1}A')$ for $A' \in \mathcal{F}'$. Prove that there exists a function $p(A|\omega')$ on Ω' , measurable \mathcal{F}' and integrable PT^{-1} , such that $\int_A p(A|\omega') PT^{-1}(d\omega') = P(A \cap T^{-1}A')$ for all A' in \mathcal{F}' . Intuitively, $p(A|\omega')$ is the conditional probability that $\omega \in A$ for someone who knows that $T\omega = \omega'$. Let $\mathcal{G} = [T^{-1}A': A' \in \mathcal{F}']$; show that \mathcal{G} is a σ -field and that $p(A|T\omega)$ is a version of $P[A|\mathcal{G}]_\omega$.
- (b) Connect this with part (a) of the preceding problem.
- 33.6. \uparrow Suppose that $T = X$ is a random variable, $(\Omega', \mathcal{F}') = (R^1, \mathcal{B}^1)$, and x is the general point of R^1 . In this case $p(A|x)$ is sometimes written $P[A|X = x]$. What is the problem with this notation?

- 33.7. For the Poisson process (see Example 33.1) show that for $0 < s < t$,

$$P[N_s = k | N_t] = \begin{cases} \binom{N_t}{k} \left(\frac{s}{t}\right)^k \left(1 - \frac{s}{t}\right)^{N_t - k}, & k \leq N_t, \\ 0, & k > N_t. \end{cases}$$

Thus the conditional distribution (in the sense of Theorem 33.3) of N_s given N_t is binomial with parameters N_t and s/t .

- 33.8. 29.12 \uparrow Suppose that (X_1, X_2) has the centered normal distribution—has in the plane the distribution with density (29.10). Express the quadratic form in the exponential as

$$\frac{1}{\sigma_{11}} x_1^2 + \frac{\sigma_{11}}{D} \left(x_2 - \frac{\sigma_{12}}{\sigma_{11}} x_1 \right)^2;$$

integrate out the x_2 and show that

$$\frac{f(x_1, x_2)}{\int_{-\infty}^{\infty} f(x_1, t) dt} = \frac{1}{\sqrt{2\pi\tau}} \exp\left[-\frac{1}{2\tau}\left(x_2 - \frac{\sigma_{12}}{\sigma_{11}}x_1\right)^2\right],$$

where $\tau = \sigma_{22} - \sigma_{12}^2\sigma_{11}^{-1}$. Describe the conditional distribution of X_2 given X_1 .

33.9. (a) Suppose that $\mu(H, \omega)$ has property (i) in Theorem 33.3, and suppose that $\mu(H, \cdot)$ is a version of $P[X \in H | \mathcal{G}]$ for H in a π -system generating \mathcal{R}^1 . Show that $\mu(\cdot, \omega)$ is a conditional distribution of X given \mathcal{G} .

(b) Use Theorem 12.5 to extend Theorem 33.3 from R^1 to R^k .

(c) Show that conditional probabilities can be defined as genuine probabilities on spaces of the special form $(\Omega, \sigma(X_1, \dots, X_k), P)$.

33.10. \uparrow Deduce from (33.16) that the conditional distribution of X given M is

$$\frac{1}{2}I_{[M \in H]}(\omega) + \frac{1}{2} \frac{\mu(H \cap (-\infty, M(\omega)]}{\mu(-\infty, M(\omega)]},$$

where μ is the distribution corresponding to F (positive and continuous). *Hint:* First check $H = (-\infty, x]$.

33.11. 4.10 12.4 \uparrow The following construction shows that conditional probabilities may not give measures. Complete the details.

In Problem 4.10 it is shown that there exist a probability space (Ω, \mathcal{F}, P) , a σ -field \mathcal{G} in \mathcal{F} , and a set H in \mathcal{F} such that $P(H) = \frac{1}{2}$, H and \mathcal{G} are independent, \mathcal{G} contains all the singletons, and \mathcal{G} is generated by a countable subclass. The countable subclass generating \mathcal{G} can be taken to be a π -system $\mathcal{P} = \{B_1, B_2, \dots\}$ (pass to the finite intersections of the sets in the original class).

Assume that it is possible to choose versions $P[A | \mathcal{G}]$ so that $P[A | \mathcal{G}]_\omega$ is for each ω a probability measure as A varies over \mathcal{F} . Let C_n be the ω -set where $P[B_n | \mathcal{G}]_\omega = I_{B_n}(\omega)$; show (Example 33.3) that $C = \bigcap_n C_n$ has probability 1. Show that $\omega \in C$ implies that $P[G | \mathcal{G}]_\omega = I_G(\omega)$ for all G in \mathcal{G} and hence that $P[\{\omega\} | \mathcal{G}]_\omega = 1$.

Now $\omega \in H \cap C$ implies that $P[H | \mathcal{G}]_\omega \geq P[\{\omega\} | \mathcal{G}]_\omega = 1$ and $\omega \in H^c \cap C$ implies that $P[H | \mathcal{G}]_\omega \leq P[\Omega - \{\omega\} | \mathcal{G}]_\omega = 0$. Thus $\omega \in C$ implies that $P[H | \mathcal{G}]_\omega = I_H(\omega)$. But since H and \mathcal{G} are independent, $P[H | \mathcal{G}]_\omega = P(H) = \frac{1}{2}$ with probability 1, a contradiction.

This example is related to Example 4.10 but concerns mathematical fact instead of heuristic interpretation.

33.12. Let α and β be σ -finite measures on the line, and let $f(x, y)$ be a probability density with respect to $\alpha \times \beta$. Define

$$(33.34) \quad g_x(y) = \frac{f(x, y)}{\int_{R^1} f(x, t) \beta(dt)},$$

unless the denominator vanishes, in which case take $g_x(y) = 0$, say. Show that, if (X, Y) has density f with respect to $\alpha \times \beta$, then the conditional distribution of Y given X has density $g_x(y)$ with respect to β . This generalizes Examples 33.5 and 33.12, where α and β are Lebesgue measure.

- 33.13.** 18.20 \uparrow Suppose that μ and ν_x (one for each real x) are probability measures on the line, and suppose that $\nu_x(B)$ is a Borel function in x for each $B \in \mathcal{R}^1$. Then (see Problem 18.20)

$$(33.35) \quad \pi(E) = \int_{R^1} \nu_x[y: (x, y) \in E] \mu(dx)$$

defines a probability measure on (R^2, \mathcal{R}^2) .

Suppose that (X, Y) has distribution π , and show that ν_x is a version of the conditional distribution of Y given X .

- 33.14.** \uparrow Let α and β be σ -finite measures on the line. Specialize the setup of Problem 33.13 by supposing that μ has density $f(x)$ with respect to α and ν_x has density $g_x(y)$ with respect to β . Assume that $g_x(y)$ is measurable \mathcal{R}^2 in the pair (x, y) , so that $\nu_x(B)$ is automatically measurable in x . Show that (33.35) has density $f(x)g_x(y)$ with respect to $\alpha \times \beta$: $\pi(E) = \iint_E f(x)g_x(y)\alpha(dx)\beta(dy)$. Show that (33.34) is consistent with $f(x, y) = f(x)g_x(y)$. Put

$$p_y(x) = \frac{f(x)g_x(y)}{\int_{R^1} f(s)g_s(y)\alpha(ds)}.$$

Suppose that (X, Y) has density $f(x)g_x(y)$ with respect to $\alpha \times \beta$, and show that $p_y(x)$ is a density with respect to α for the conditional distribution of X given Y .

In the language of Bayes, $f(x)$ is the prior density of a parameter x , $g_x(y)$ is the conditional density of the observation y given the parameter, and $p_y(x)$ is the posterior density of the parameter given the observation.

- 33.15.** \uparrow Now suppose that α and β are Lebesgue measure, that $f(x)$ is positive, continuous, and bounded, and that $g_x(y) = e^{-(y-x)^2 n/2} / \sqrt{2\pi/n}$. Thus the observation is distributed as the average of n independent normal variables with mean x and variance 1. Show that

$$\frac{1}{\sqrt{n}} p_y\left(y + \frac{x}{\sqrt{n}}\right) \rightarrow \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$$

for fixed x and y . Thus the posterior density is approximately that of a normal distribution with mean y and variance $1/n$.

- 33.16.** 32.13 \uparrow Suppose that X has distribution μ . Now $P[A|X]_\omega = f(X(\omega))$ for some Borel function f . Show that $\lim_{h \rightarrow 0} P[A|x-h < X \leq x+h] = f(x)$ for x in a set of μ -measure 1. Roughly speaking, $P[A|x-h < X \leq x+h] \rightarrow P[A|X=x]$. *Hint:* Take $\nu(B) = P(A \cap [X \in B])$ in Problem 32.13.

SECTION 34. CONDITIONAL EXPECTATION

In this section the theory of conditional expectation is developed from first principles. The properties of conditional probabilities will then follow as special cases. The preceding section was long only because of the examples in it; the theory itself is quite compact.

Definition

Suppose that X is an integrable random variable on (Ω, \mathcal{F}, P) and that \mathcal{G} is a σ -field in \mathcal{F} . There exists a random variable $E[X|\mathcal{G}]$, called the *conditional expected value* of X given \mathcal{G} , having these two properties:

- (i) $E[X|\mathcal{G}]$ is measurable \mathcal{G} and integrable.
- (ii) $E[X|\mathcal{G}]$ satisfies the functional equation

$$(34.1) \quad \int_G E[X|\mathcal{G}] dP = \int_G X dP, \quad G \in \mathcal{G}.$$

To prove the existence of such a random variable, consider first the case of nonnegative X . Define a measure ν on \mathcal{G} by $\nu(G) = \int_G X dP$. This measure is finite because X is integrable, and it is absolutely continuous with respect to P . By the Radon–Nikodym theorem there is a function f , measurable \mathcal{G} , such that $\nu(G) = \int_G f dP$.[†] This f has properties (i) and (ii). If X is not necessarily nonnegative, $E[X^+|\mathcal{G}] - E[X^-|\mathcal{G}]$ clearly has the required properties.

There will in general be many such random variables $E[X|\mathcal{G}]$; any one of them is called a *version* of the conditional expected value. Any two versions are equal with probability 1 (by Theorem 16.10 applied to P restricted to \mathcal{G}).

Arguments like those in Examples 33.3 and 33.4 show that $E[X|\{0, \Omega\}] = E[X]$ and that $E[X|\mathcal{F}] = X$ with probability 1. As \mathcal{G} increases, condition (i) becomes weaker and condition (ii) becomes stronger.

The value $E[X|\mathcal{G}]_\omega$ at ω is to be interpreted as the expected value of X for someone who knows for each G in \mathcal{G} whether or not it contains the point ω , which itself in general remains unknown. Condition (i) ensures that $E[X|\mathcal{G}]$ can in principle be calculated from this partial information alone. Condition (ii) can be restated as $\int_G (X - E[X|\mathcal{G}]) dP = 0$; if the observer, in possession of the partial information contained in \mathcal{G} , is offered the opportunity to bet, paying an entry fee of $E[X|\mathcal{G}]$ and being returned the amount X , and if he adopts the strategy of betting if G occurs, this equation says that the game is fair.

[†]As in the case of conditional probabilities, the integral is the same on (Ω, \mathcal{F}, P) as on (Ω, \mathcal{G}) with P restricted to \mathcal{G} (Example 16.4).

Example 34.1. Suppose that B_1, B_2, \dots is a finite or countable partition of Ω generating the σ -field \mathcal{G} . Then $E[X|\mathcal{G}]$ must, since it is measurable \mathcal{G} , have some constant value over B_i , say α_i . Then (34.1) for $G = B_i$ gives $\alpha_i P(B_i) = \int_{B_i} X dP$. Thus

$$(34.2) \quad E[X|\mathcal{G}]_\omega = \frac{1}{P(B_i)} \int_{B_i} X dP, \quad \omega \in B_i, \quad P(B_i) > 0.$$

If $P(B_i) = 0$, the value of $E[X|\mathcal{G}]$ over B_i is constant but arbitrary. ■

Example 34.2. For an indicator I_A the defining properties of $E[I_A|\mathcal{G}]$ and $P[A|\mathcal{G}]$ coincide; therefore, $E[I_A|\mathcal{G}] = P[A|\mathcal{G}]$ with probability 1. It is easily checked that, more generally, $E[X|\mathcal{G}] = \sum_i \alpha_i P[A_i|\mathcal{G}]$ with probability 1 for a simple function $X = \sum_i \alpha_i I_{A_i}$. ■

In analogy with the case of conditional probability, if $[X_t, t \in T]$ is a collection of random variables, $E[X|X_t, t \in T]$ is by definition $E[X|\mathcal{G}]$ with $\sigma[X_t, t \in T]$ in the role of \mathcal{G} .

Example 34.3. Let \mathcal{I} be the σ -field of sets invariant under a measure-preserving transformation T on (Ω, \mathcal{F}, P) . For f integrable, the limit \hat{f} in (24.7) is $E[f|\mathcal{I}]$: Since \hat{f} is invariant, it is measurable \mathcal{I} . If G is invariant, then the averages a_n in the proof of the ergodic theorem (p. 318) satisfy $E[I_G a_n] = E[I_G f]$. But since the a_n converge to \hat{f} and are uniformly integrable, $E[I_G \hat{f}] = E[I_G f]$. ■

Properties of Conditional Expectation

To prove the first result, apply Theorem 16.10(iii) to f and $E[X|\mathcal{G}]$ on (Ω, \mathcal{G}, P) .

Theorem 34.1. Let \mathcal{P} be a π -system generating the σ -field \mathcal{G} , and suppose that Ω is a finite or countable union of sets in \mathcal{G} . An integrable function f is a version of $E[X|\mathcal{G}]$ if it is measurable \mathcal{G} and if

$$(34.3) \quad \int_G f dP = \int_G X dP$$

holds for all G in \mathcal{P} .

In most applications it is clear that $\Omega \in \mathcal{P}$.

All the equalities and inequalities in the following theorem hold with probability 1.

Theorem 34.2. *Suppose that X, Y, X_n are integrable.*

- (i) *If $X = a$ with probability 1, then $E[X|\mathcal{G}] = a$.*
- (ii) *For constants a and b , $E[aX + bY|\mathcal{G}] = aE[X|\mathcal{G}] + bE[Y|\mathcal{G}]$.*
- (iii) *If $X \leq Y$ with probability 1, then $E[X|\mathcal{G}] \leq E[Y|\mathcal{G}]$.*
- (iv) *$|E[X|\mathcal{G}]| \leq E[|X||\mathcal{G}]$.*
- (v) *If $\lim_n X_n = X$ with probability 1, $|X_n| \leq Y$, and Y is integrable, then $\lim_n E[X_n|\mathcal{G}] = E[X|\mathcal{G}]$ with probability 1.*

PROOF. If $X = a$ with probability 1, the function identically equal to a satisfies conditions (i) and (ii) in the definition of $E[X|\mathcal{G}]$, and so (i) above follows by uniqueness.

As for (ii), $aE[X|\mathcal{G}] + bE[Y|\mathcal{G}]$ is integrable and measurable \mathcal{G} , and

$$\begin{aligned} \int_G (aE[X|\mathcal{G}] + bE[Y|\mathcal{G}]) dP &= a \int_G E[X|\mathcal{G}] dP + b \int_G E[Y|\mathcal{G}] dP \\ &= a \int_G X dP + b \int_G Y dP = \int_G (aX + bY) dP \end{aligned}$$

for all G in \mathcal{G} , so that this function satisfies the functional equation.

If $X \leq Y$ with probability 1, then $\int_G (E[Y|\mathcal{G}] - E[X|\mathcal{G}]) dP = \int_G (Y - X) dP \geq 0$ for all G in \mathcal{G} . Since $E[Y|\mathcal{G}] - E[X|\mathcal{G}]$ is measurable \mathcal{G} , it must be nonnegative with probability 1 (consider the set G where it is negative). This proves (iii), which clearly implies (iv) as well as the fact that $E[X|\mathcal{G}] = E[Y|\mathcal{G}]$ if $X = Y$ with probability 1.

To prove (v), consider $Z_n = \sup_{k \geq n} |X_k - X|$. Now $Z_n \downarrow 0$ with probability 1, and by (ii), (iii), and (iv), $|E[X_n|\mathcal{G}] - E[X|\mathcal{G}]| \leq E[Z_n|\mathcal{G}]$. It suffices, therefore, to show that $E[Z_n|\mathcal{G}] \downarrow 0$ with probability 1. By (iii) the sequence $E[Z_n|\mathcal{G}]$ is nonincreasing and hence has a limit Z ; the problem is to prove that $Z = 0$ with probability 1, or, Z being nonnegative, that $E[Z] = 0$. But $0 \leq Z_n \leq 2Y$, and so (34.1) and the dominated convergence theorem give $E[Z] = \int E[Z|\mathcal{G}] dP \leq \int E[Z_n|\mathcal{G}] dP = E[Z_n] \rightarrow 0$. ■

The properties (33.21) through (33.28) can be derived anew from Theorem 34.2. Part (ii) shows once again that $E[\sum_i \alpha_i I_{A_i}|\mathcal{G}] = \sum_i \alpha_i P[A_i|\mathcal{G}]$ for simple functions.

If X is measurable \mathcal{G} , then clearly $E[X|\mathcal{G}] = X$ with probability 1. The following generalization of this is used constantly. For an observer with the information in \mathcal{G} , X is effectively a constant if it is measurable \mathcal{G} :

Theorem 34.3. *If X is measurable \mathcal{G} , and if Y and XY are integrable, then*

$$(34.4) \quad E[XY|\mathcal{G}] = XE[Y|\mathcal{G}]$$

with probability 1.

PROOF. It will be shown first that the right side of (34.4) is a version of the left side if $X = I_{G_0}$ and $G_0 \in \mathcal{G}$. Since $I_{G_0}E[Y|\mathcal{G}]$ is certainly measurable \mathcal{G} , it suffices to show that it satisfies the functional equation $\int_G I_{G_0}E[Y|\mathcal{G}]dP = \int_G I_{G_0}YdP$, $G \in \mathcal{G}$. But this reduces to $\int_{G \cap G_0} E[Y|\mathcal{G}]dP = \int_{G \cap G_0} YdP$, which holds by the definition of $E[Y|\mathcal{G}]$. Thus (34.4) holds if X is the indicator of an element of \mathcal{G} .

It follows by Theorem 34.2(ii) that (34.4) holds if X is a simple function measurable \mathcal{G} . For the general X that is measurable \mathcal{G} , there exist simple functions X_n , measurable \mathcal{G} , such that $|X_n| \leq |X|$ and $\lim_n X_n = X$ (Theorem 13.5). Since $|X_n Y| \leq |XY|$ and $|XY|$ is integrable, Theorem 34.2(v) implies that $\lim_n E[X_n Y|\mathcal{G}] = E[XY|\mathcal{G}]$ with probability 1. But $E[X_n Y|\mathcal{G}] = X_n E[Y|\mathcal{G}]$ by the case already treated, and of course $\lim_n X_n E[Y|\mathcal{G}] = XE[Y|\mathcal{G}]$. (Note that $|X_n E[Y|\mathcal{G}]| = |E[X_n Y|\mathcal{G}]| \leq E[|X_n Y||\mathcal{G}] \leq E[|XY||\mathcal{G}]$, so that the limit $XE[Y|\mathcal{G}]$ is integrable.) Thus (34.4) holds in general. Notice that X has not been assumed integrable. ■

Taking a conditional expected value can be thought of as an averaging or smoothing operation. This leads one to expect that averaging X with respect to \mathcal{G}_2 and then averaging the result with respect to a coarser (smaller) σ -field \mathcal{G}_1 should lead to the same result as would averaging with respect to \mathcal{G}_1 in the first place:

Theorem 34.4. *If X is integrable and the σ -fields \mathcal{G}_1 and \mathcal{G}_2 satisfy $\mathcal{G}_1 \subset \mathcal{G}_2$, then*

$$(34.5) \quad E[E[X|\mathcal{G}_2]|\mathcal{G}_1] = E[X|\mathcal{G}_1]$$

with probability 1.

PROOF. The left side of (34.5) is measurable \mathcal{G}_1 , and so to prove that it is a version of $E[X|\mathcal{G}_1]$, it is enough to verify $\int_G E[E[X|\mathcal{G}_2]|\mathcal{G}_1]dP = \int_G XdP$ for $G \in \mathcal{G}_1$. But if $G \in \mathcal{G}_1$, then $G \in \mathcal{G}_2$, and the left side here is $\int_G E[X|\mathcal{G}_2]dP = \int_G XdP$. ■

If $\mathcal{G}_2 = \mathcal{F}$, then $E[X|\mathcal{G}_2] = X$, so that (34.5) is trivial. If $\mathcal{G}_1 = \{0, \Omega\}$ and $\mathcal{G}_2 = \mathcal{G}$, then (34.5) becomes

$$(34.6) \quad E[E[X|\mathcal{G}]] = E[X],$$

the special case of (34.1) for $G = \Omega$.

If $\mathcal{G}_1 \subset \mathcal{G}_2$, then $E[X|\mathcal{G}_1]$, being measurable \mathcal{G}_1 , is also measurable \mathcal{G}_2 , so that taking an expected value with respect to \mathcal{G}_2 does not alter it: $E[E[X|\mathcal{G}_1]|\mathcal{G}_2] = E[X|\mathcal{G}_1]$. Therefore, if $\mathcal{G}_1 \subset \mathcal{G}_2$, taking iterated expected values in either order gives $E[X|\mathcal{G}_1]$.

The remaining result of a general sort needed here is *Jensen's inequality* for conditional expected values: If φ is a convex function on the line and X and $\varphi(X)$ are both integrable, then

$$(34.7) \quad \varphi(E[X|\mathcal{G}]) \leq E[\varphi(X)|\mathcal{G}]$$

with probability 1. For each x_0 take a support line [A33] through $(x_0, \varphi(x_0))$: $\varphi(x_0) + A(x_0)(x - x_0) \leq \varphi(x)$. The slope $A(x_0)$ can be taken as the right-hand derivative of φ , so that it is nondecreasing in x_0 . Now

$$\varphi(E[X|\mathcal{G}]) + A(E[X|\mathcal{G}])(X - E[X|\mathcal{G}]) \leq \varphi(X).$$

Suppose that $E[X|\mathcal{G}]$ is bounded. Then all three terms here are integrable (if φ is convex on R^1 , then φ and A are bounded on bounded sets), and taking expected values with respect to \mathcal{G} and using (34.4) on the middle term gives (34.7).

To prove (34.7) in general, let $G_n = [|E[X|\mathcal{G}]| \leq n]$. Then $E[I_{G_n}X|\mathcal{G}] = I_{G_n}E[X|\mathcal{G}]$ is bounded, and so (34.7) holds for $I_{G_n}X$: $\varphi(I_{G_n}E[X|\mathcal{G}]) \leq E[\varphi(I_{G_n}X)|\mathcal{G}]$. Now $E[\varphi(I_{G_n}X)|\mathcal{G}] = E[I_{G_n}\varphi(X) + I_{G_n^c}\varphi(0)|\mathcal{G}] = I_{G_n}E[\varphi(X)|\mathcal{G}] + I_{G_n^c}\varphi(0) \rightarrow E[\varphi(X)|\mathcal{G}]$. Since $\varphi(I_{G_n}E[X|\mathcal{G}])$ converges to $\varphi(E[X|\mathcal{G}])$ by the continuity of φ , (34.7) follows. If $\varphi(x) = |x|$, (34.7) gives part (iv) of Theorem 34.2 again.

Conditional Distributions and Expectations

Theorem 34.5. *Let $\mu(\cdot, \omega)$ be a conditional distribution with respect to \mathcal{G} of a random variable X , in the sense of Theorem 33.3. If $\varphi: R^1 \rightarrow R^1$ is a Borel function for which $\varphi(X)$ is integrable, then $\int_{R^1} \varphi(x) \mu(dx, \omega)$ is a version of $E[\varphi(X)|\mathcal{G}]_\omega$.*

PROOF. If $\varphi = I_H$ and $H \in \mathcal{R}^1$, this is an immediate consequence of the definition of conditional distribution, and by Theorem 34.2(ii) it follows for φ a simple function over R^1 . For the general nonnegative φ , choose simple φ_n such that $0 \leq \varphi_n(x) \uparrow \varphi(x)$ for each x in R^1 . By the case already treated, $\int_{R^1} \varphi_n(x) \mu(dx, \omega)$ is a version of $E[\varphi_n(X)|\mathcal{G}]_\omega$. The integral converges by the monotone convergence theorem in $(R^1, \mathcal{R}^1, \mu(\cdot, \omega))$ to $\int_{R^1} \varphi(x) \mu(dx, \omega)$ for each ω , the value $+\infty$ not excluded, and $E[\varphi_n(X)|\mathcal{G}]_\omega$ converges to $E[\varphi(X)|\mathcal{G}]_\omega$ with probability 1 by Theorem 34.2(v). Thus the result holds for nonnegative φ , and the general case follows from splitting into positive and negative parts. ■

It is a consequence of the proof above that $\int_{R^1} \varphi(x) \mu(dx, \omega)$ is measurable \mathcal{G} and finite with probability 1. If X is itself integrable, it follows by the

theorem for the case $\varphi(x) = x$ that

$$E[X|\mathcal{G}]_\omega = \int_{-\infty}^{\infty} x \mu(dx, \omega)$$

with probability 1. If $\varphi(X)$ is integrable as well, then

$$(34.8) \quad E[\varphi(X)|\mathcal{G}]_\omega = \int_{-\infty}^{\infty} \varphi(x) \mu(dx, \omega)$$

with probability 1. By Jensen's inequality (21.14) for unconditional expected values, the right side of (34.8) is at least $\varphi(\int_{-\infty}^{\infty} x \mu(dx, \omega))$ if φ is convex. This gives another proof of (34.7).

Sufficient Subfields*

Suppose that for each θ in an index set Θ , P_θ is a probability measure on (Ω, \mathcal{F}) . In statistics the problem is to draw inferences about the unknown parameter θ from an observation ω .

Denote by $P_\theta[A|\mathcal{G}]$ and $E_\theta[X|\mathcal{G}]$ conditional probabilities and expected values calculated with respect to the probability measure P_θ on (Ω, \mathcal{F}) . A σ -field \mathcal{G} in \mathcal{F} is *sufficient* for the family $[P_\theta: \theta \in \Theta]$ if versions $P_\theta[A|\mathcal{G}]$ can be chosen that are independent of θ —that is, if there exists a function $p(A, \omega)$ of $A \in \mathcal{F}$ and $\omega \in \Omega$ such that, for each $A \in \mathcal{F}$ and $\theta \in \Theta$, $p(A, \cdot)$ is a version of $P_\theta[A|\mathcal{G}]$. There is no requirement that $p(\cdot, \omega)$ be a measure for ω fixed. The idea is that although there may be information in \mathcal{F} not already contained in \mathcal{G} , this information is irrelevant to the drawing of inferences about θ .[†] A *sufficient statistic* is a random variable or random vector T such that $\sigma(T)$ is a sufficient subfield.

A family \mathcal{M} of measures *dominates* another family \mathcal{N} if, for each A , from $\mu(A) = 0$ for all μ in \mathcal{M} , it follows that $\nu(A) = 0$ for all ν in \mathcal{N} . If each of \mathcal{M} and \mathcal{N} dominates the other, they are *equivalent*. For sets consisting of a single measure these are the concepts introduced in Section 32.

Theorem 34.6. *Suppose that $[P_\theta: \theta \in \Theta]$ is dominated by the σ -finite measure μ . A necessary and sufficient condition for \mathcal{G} to be sufficient is that the density f_θ of P_θ with respect to μ can be put in the form $f_\theta = g_\theta h$ for a g_θ measurable \mathcal{G} .*

It is assumed throughout that g_θ and h are nonnegative and of course that h is measurable \mathcal{F} . Theorem 34.6 is called the *factorization theorem*, the condition being that the density f_θ splits into a factor depending on ω only through \mathcal{G} and a factor independent of θ . Although g_θ and h are not assumed integrable μ , their product f_θ , as the density of a finite measure, must be. Before proceeding to the proof, consider an application.

Example 34.4. Let $(\Omega, \mathcal{F}) = (R^k, \mathcal{R}^k)$, and for $\theta > 0$ let P_θ be the measure having with respect to k -dimensional Lebesgue measure the density

$$f_\theta(x) = f_\theta(x_1, \dots, x_k) = \begin{cases} \theta^{-k} & \text{if } 0 \leq x_i \leq \theta, i = 1, \dots, k, \\ 0 & \text{otherwise.} \end{cases}$$

*This topic may be omitted.

[†]See Problem 34.19.

If X_i is the function on R^k defined by $X_i(x) = x_i$, then under P_θ , X_1, \dots, X_k are independent random variables, each uniformly distributed over $[0, \theta]$. Let $T(x) = \max_{i \leq k} X_i(x)$. If $g_\theta(t)$ is θ^{-k} for $0 \leq t \leq \theta$ and 0 otherwise, and if $h(x)$ is 1 or 0 according as all x_i are nonnegative or not, then $f_\theta(x) = g_\theta(T(x))h(x)$. The factorization criterion is thus satisfied, and T is a sufficient statistic.

Sufficiency is clear on intuitive grounds as well: θ is not involved in the conditional distribution of X_1, \dots, X_k given T because, roughly speaking, a random one of them equals T and the others are independent and uniform over $[0, T]$. If this is true, the distribution of X_i given T ought to have a mass of k^{-1} at T and a uniform distribution of mass $1 - k^{-1}$ over $[0, T]$, so that

$$(34.9) \quad E_\theta[X_i | T] = \frac{1}{k} T + \frac{k-1}{k} \frac{T}{2} = \frac{k+1}{2k} T.$$

For a proof of this fact, needed later, note that by (21.9)

$$(34.10) \quad \begin{aligned} \int_{T \leq t} X_i dP_\theta &= \int_0^\infty P_\theta[T \leq t, X_i \geq u] du \\ &= \int_0^t \frac{t-u}{\theta} \left(\frac{t}{\theta}\right)^{k-1} du = \frac{t^{k+1}}{2\theta^k} \end{aligned}$$

if $0 \leq t \leq \theta$. On the other hand, $P_\theta[T \leq t] = (t/\theta)^k$, so that under P_θ the distribution of T has density kt^{k-1}/θ^k over $[0, \theta]$. Thus

$$(34.11) \quad \int_{T \leq t} \frac{k+1}{2k} T dP_\theta = \frac{k+1}{2k} \int_0^t uk \frac{u^{k-1}}{\theta^k} du = \frac{t^{k+1}}{2\theta^k}.$$

Since (34.10) and (34.11) agree, (34.9) follows by Theorem 34.1. ■

The essential ideas in the proof of Theorem 34.6 are most easily understood through a preliminary consideration of special cases.

Lemma 1. Suppose that $[P_\theta: \theta \in \Theta]$ is dominated by a probability measure P and that each P_θ has with respect to P a density g_θ that is measurable \mathcal{G} . Then \mathcal{G} is sufficient, and $P[A|\mathcal{G}]$ is a version of $P_\theta[A|\mathcal{G}]$ for each θ in Θ .

PROOF. For G in \mathcal{G} , (34.4) gives

$$\begin{aligned} \int_G P[A|\mathcal{G}] dP_\theta &= \int_G E[I_A|\mathcal{G}] g_\theta dP = \int_G E[I_A g_\theta|\mathcal{G}] dP \\ &= \int_G I_A g_\theta dP = \int_{A \cap G} g_\theta dP = P_\theta(A \cap G). \end{aligned}$$

Therefore, $P[A|\mathcal{G}]$ —the conditional probability calculated with respect to P —does serve as a version of $P_\theta[A|\mathcal{G}]$ for each θ in Θ . Thus \mathcal{G} is sufficient for the family

$[P_\theta: \theta \in \Theta]$ —even for this family augmented by P (which might happen to lie in the family to start with). ■

For the necessity, suppose first that the family is dominated by one of its members.

Lemma 2. *Suppose that $[P_\theta: \theta \in \Theta]$ is dominated by P_{θ_0} for some $\theta_0 \in \Theta$. If \mathcal{S} is sufficient, then each P_θ has with respect to P_{θ_0} a density g_θ that is measurable \mathcal{S} .*

PROOF. Let $p(A, \omega)$ be the function in the definition of sufficiency, and take $P_\theta[A|\mathcal{S}]_\omega = p(A, \omega)$ for all $A \in \mathcal{F}$, $\omega \in \Omega$, and $\theta \in \Theta$. Let d_θ be any density of P_θ with respect to P_{θ_0} . By a number of applications of (34.4),

$$\begin{aligned} \int_A E_{\theta_0}[d_\theta|\mathcal{S}] dP_{\theta_0} &= \int I_A E_{\theta_0}[d_\theta|\mathcal{S}] dP_{\theta_0} \\ &= \int E_{\theta_0}\{I_A E_{\theta_0}[d_\theta|\mathcal{S}]\} dP_{\theta_0} = \int E_{\theta_0}\{I_A|\mathcal{S}\} E_{\theta_0}[d_\theta|\mathcal{S}] dP_{\theta_0} \\ &= \int E_{\theta_0}[E_{\theta_0}\{I_A|\mathcal{S}\} d_\theta|\mathcal{S}] dP_{\theta_0} = \int E_{\theta_0}\{I_A|\mathcal{S}\} d_\theta dP_{\theta_0} \\ &= \int P_{\theta_0}[A|\mathcal{S}] dP_\theta = \int P_\theta[A|\mathcal{S}] dP_\theta = P_\theta(A), \end{aligned}$$

the next-to-last equality by sufficiency (the integrand on either side being $p(A, \cdot)$). Thus $g_\theta = E_{\theta_0}[d_\theta|\mathcal{S}]$, which is measurable \mathcal{S} , can serve as a density for P_θ with respect to P_{θ_0} . ■

To complete the proof of Theorem 34.6 requires one more lemma of a technical sort.

Lemma 3. *If $[P_\theta: \theta \in \Theta]$ is dominated by a σ -finite measure, then it is equivalent to some finite or countably infinite subfamily.*

In many examples, the P_θ are all equivalent to each other, in which case the subfamily can be taken to consist of a single P_{θ_0} .

PROOF. Since μ is σ -finite, there is a finite or countable partition of Ω into \mathcal{F} -sets A_n such that $0 < \mu(A_n) < \infty$. Choose positive constants a_n , one for each A_n , in such a way that $\sum_n a_n < \infty$. The finite measure with value $\sum_n a_n \mu(A \cap A_n) / \mu(A_n)$ at A dominates μ . In proving the lemma it is therefore no restriction to assume the family dominated by a *finite* measure μ .

Each P_θ is dominated by μ and hence has a density f_θ with respect to it. Let $S_\theta = [\omega: f_\theta(\omega) > 0]$. Then $P_\theta(A) = P_\theta(A \cap S_\theta)$ for all A , and $P_\theta(A) = 0$ if and only if $\mu(A \cap S_\theta) = 0$. In particular, S_θ supports P_θ .

Call a set B in \mathcal{F} a *kernel* if $B \subset S_\theta$ for some θ , and call a finite or countable union of kernels a *chain*. Let α be the supremum of $\mu(C)$ over chains C . Since μ is finite and a finite or countable union of chains is a chain, α is finite and $\mu(C) = \alpha$ for some chain C . Suppose that $C = \bigcup_n B_n$, where each B_n is a kernel, and suppose that $B_n \subset S_{\theta_n}$.

The problem is to show that $[P_\theta: \theta \in \Theta]$ is dominated by $[P_{\theta_n}: n = 1, 2, \dots]$ and hence equivalent to it. Suppose that $P_{\theta_n}(A) = 0$ for all n . Then $\mu(A \cap S_{\theta_n}) = 0$, as observed above. Since $C \subset \bigcup_n S_{\theta_n}$, $\mu(A \cap C) = 0$, and it follows that $P_\theta(A \cap C) = 0$

whatever θ may be. But suppose that $P_\theta(A - C) > 0$. Then $P_\theta((A - C) \cap S_\theta) = P_\theta(A - C)$ is positive, and so $(A - C) \cap S_\theta$ is a kernel, disjoint from C , of positive μ -measure; this is impossible because of the maximality of C . Thus $P_\theta(A - C)$ is 0 along with $P_\theta(A \cap C)$, and so $P_\theta(A) = 0$. ■

Suppose that $[P_\theta: \theta \in \Theta]$ is dominated by a σ -finite μ , as in Theorem 34.6, so that, by Lemma 3, it contains a finite or infinite sequence $P_{\theta_1}, P_{\theta_2}, \dots$ equivalent to the entire family. Fix one such sequence, and choose positive constants c_n , one for each θ_n , in such a way that $\sum_n c_n = 1$. Now define a probability measure P on \mathcal{F} by

$$(34.12) \quad P(A) = \sum_n c_n P_{\theta_n}(A).$$

Clearly, P is equivalent to $[P_{\theta_1}, P_{\theta_2}, \dots]$ and hence to $[P_\theta: \theta \in \Theta]$, and all three are dominated by μ :

$$(34.13) \quad P \equiv [P_{\theta_1}, P_{\theta_2}, \dots] \equiv [P_\theta: \theta \in \Theta] \ll \mu.$$

PROOF OF SUFFICIENCY IN THEOREM 34.6. If each P_θ has density $g_\theta h$ with respect to μ , then by the construction (34.12), P has density fh with respect to μ , where $f = \sum_n c_n g_{\theta_n}$. Put $r_\theta = g_\theta/f$ if $f > 0$, and $r_\theta = 0$ (say) if $f = 0$. If each g_θ is measurable \mathcal{G} , the same is true of f and hence of the r_θ . Since $P[f = 0] = 0$ and P is equivalent to the entire family, $P_\theta[f = 0] = 0$ for all θ . Therefore,

$$\begin{aligned} \int_A r_\theta dP &= \int_A r_\theta fh d\mu = \int_{A \cap [f > 0]} r_\theta fh d\mu = \int_{A \cap [f > 0]} g_\theta h d\mu \\ &= P_\theta(A \cap [f > 0]) = P_\theta(A). \end{aligned}$$

Each P_θ thus has with respect to the probability measure P a density measurable \mathcal{G} , and it follows by Lemma 1 that \mathcal{G} is sufficient. ■

PROOF OF NECESSITY IN THEOREM 34.6. Let $p(A, \omega)$ be a function such that, for each A and θ , $p(A, \cdot)$ is a version of $P_\theta[A|\mathcal{G}]$, as required by the definition of sufficiency. For P as in (34.12) and $G \in \mathcal{G}$,

$$\begin{aligned} (34.14) \quad \int_G p(A, \omega) P(d\omega) &= \sum_n c_n \int_G p(A, \omega) P_{\theta_n}(d\omega) \\ &= \sum_n c_n \int_G P_{\theta_n}[A|\mathcal{G}] dP_{\theta_n} \\ &= \sum_n c_n P_{\theta_n}(A \cap G) = P(A \cap G). \end{aligned}$$

Thus $p(A, \cdot)$ serves as a version of $P[A|\mathcal{G}]$ as well, and \mathcal{G} is still sufficient if P is added to the family. Since P dominates the augmented family, Lemma 2 implies that each P_θ has with respect to P a density g_θ that is measurable \mathcal{G} . But if h is the density of P with respect to μ (see (34.13)), then P_θ has density $g_\theta h$ with respect to μ . ■

A sub- σ -field \mathcal{G}_0 sufficient with respect to $[P_\theta: \theta \in \Theta]$ is *minimal* if, for each sufficient \mathcal{G} , \mathcal{G}_0 is essentially contained in \mathcal{G} in the sense that for each A in \mathcal{G}_0 there is a B in \mathcal{G} such that $P_\theta(A \Delta B) = 0$ for all θ in Θ . A sufficient \mathcal{G} represents a compression of the information in \mathcal{F} , and a minimal sufficient \mathcal{G}_0 represents the greatest possible compression.

Suppose the densities f_θ of the P_θ with respect to μ have the property that $f_\theta(\omega)$ is measurable $\mathcal{E} \times \mathcal{F}$, where \mathcal{E} is a σ -field in Θ . Let π be a probability measure on \mathcal{E} , and define P as $\int_\Theta P_\theta \pi(d\theta)$, in the sense that $P(A) = \int_\Theta \int_A f_\theta(\omega) \mu(d\omega) \pi(d\theta) = \int_\Theta P_\theta(A) \pi(d\theta)$. Obviously, $P \ll [P_\theta: \theta \in \Theta]$. Assume that

$$(34.15) \quad [P_\theta: \theta \in \Theta] \ll P = \int_\Theta P_\theta \pi(d\theta).$$

If π has mass c_n at θ_n , then P is given by (34.12), and of course, (35.15) holds if (34.13) does. Let r_θ be a density for P_θ with respect to P .

Theorem 34.7. *If (34.15) holds, then $\mathcal{G}_0 = \sigma[r_\theta: \theta \in \Theta]$ is a minimal sufficient sub- σ -field.*

PROOF. That \mathcal{G}_0 is sufficient follows by Theorem 34.6. Suppose that \mathcal{G} is sufficient. It follows by a simple extension of (34.14) that \mathcal{G} is still sufficient if P is added to the family, and then it follows by Lemma 2 that each P_θ has with respect to P a density g_θ that is measurable \mathcal{G} . Since densities are essentially unique, $P[g_\theta = r_\theta] = 1$. Let \mathcal{H} be the class of A in \mathcal{G}_0 such that $P(A \Delta B) = 0$ for some B in \mathcal{G} . Then \mathcal{H} is a σ -field containing each set of the form $A = [r_\theta \in H]$ (take $B = [g_\theta \in H]$) and hence containing \mathcal{G}_0 . Since, by (34.15), P dominates each P_θ , \mathcal{G}_0 is essentially contained in \mathcal{G} , in the sense of the definition. ■

Minimum-Variance Estimation*

To illustrate sufficiency, let g be a real function on Θ , and consider the problem of estimating $g(\theta)$. One possibility is that Θ is a subset of the line and g is the identity; another is that Θ is a subset of R^k and g picks out one of the coordinates. (This problem is considered from a slightly different point of view at the end of Section 19.) An *estimate* of $g(\theta)$ is a random variable Z , and the estimate is *unbiased* if $E_\theta[Z] = g(\theta)$ for all θ . One measure of the accuracy of the estimate Z is $E_\theta[(Z - g(\theta))^2]$.

If \mathcal{G} is sufficient, it follows by linearity (Theorem 34.2(ii)) that $E_\theta[X|\mathcal{G}]$ has for simple X a version that is independent of θ . Since there are simple X_n such that $|X_n| \leq |X|$ and $X_n \rightarrow X$, the same is true of any X that is integrable with respect to each P_θ (use Theorem 34.2(v)). Suppose that \mathcal{G} is, in fact, sufficient, and denote by $E[X|\mathcal{G}]$ a version of $E_\theta[X|\mathcal{G}]$ that is independent of θ .

Theorem 34.8. *Suppose that $E_\theta[(Z - g(\theta))^2] < \infty$ for all θ and that \mathcal{G} is sufficient. Then*

$$(34.16) \quad E_\theta[(E[Z|\mathcal{G}] - g(\theta))^2] \leq E_\theta[(Z - g(\theta))^2]$$

for all θ . If Z is unbiased, then so is $E[Z|\mathcal{G}]$.

*This topic may be omitted.

PROOF. By Jensen's inequality (34.7) for $\varphi(x) = (x - g(\theta))^2$, $(E[Z|\mathcal{G}] - g(\theta))^2 \leq E_\theta[(Z - g(\theta))^2|\mathcal{G}]$. Applying E_θ to each side gives (34.16). The second statement follows from the fact that $E_\theta[E[Z|\mathcal{G}]] = E_\theta[Z]$. ■

This, the *Rao-Blackwell theorem*, says that $E[Z|\mathcal{G}]$ is at least as good an estimate as Z if \mathcal{G} is sufficient.

Example 34.5. Returning to Example 34.4, note that each X_i has mean $\theta/2$ under P_θ , so that if $\bar{X} = k^{-1}\sum_{i=1}^k X_i$ is the sample mean, then $2\bar{X}$ is an unbiased estimate of θ . But there is a better one. By (34.9), $E_\theta[2\bar{X}|T] = (k+1)T/k = T'$, and by the Rao-Blackwell theorem, T' is an unbiased estimate with variance at most that of $2\bar{X}$.

In fact, for an arbitrary unbiased estimate Z , $E_\theta[(T' - \theta)^2] \leq E_\theta[(Z - \theta)^2]$. To prove this, let $\delta = T' - E[Z|T]$. By Theorem 20.1(ii), $\delta = f(T)$ for some Borel function f , and $E_\theta[f(T)] = 0$ for all θ . Taking account of the density for T leads to $\int_0^\theta f(x)x^{k-1}dx = 0$, so that $f(x)x^{k-1}$ integrates to 0 over all intervals. Therefore, $f(x)$ along with $f(x)x^{k-1}$ vanishes for $x > 0$, except on a set of Lebesgue measure 0, and hence $P_\theta[f(T) = 0] = 1$ and $P_\theta[T' = E[Z|T]] = 1$ for all θ . Therefore, $E_\theta[(T' - \theta)^2] = E_\theta[(E[Z|T] - \theta)^2] \leq E_\theta[(Z - \theta)^2]$ for Z unbiased, and T' has minimum variance among all unbiased estimates of θ . ■

PROBLEMS

- 34.1. Work out for conditional expected values the analogues of Problems 33.4, 33.5, and 33.9.
- 34.2. In the context of Examples 33.5 and 33.12, show that the conditional expected value of Y (if it is integrable) given X is $g(X)$, where

$$g(x) = \frac{\int_{-\infty}^{\infty} f(x, y)y dy}{\int_{-\infty}^{\infty} f(x, y) dy}.$$

- 34.3. Show that the independence of X and Y implies that $E[Y|X] = E[Y]$, which in turn implies that $E[XY] = E[X]E[Y]$. Show by examples in an Ω of three points that the reverse implications are both false.
- 34.4. (a) Let B be an event with $P(B) > 0$, and define a probability measure P_0 by $P_0(A) = P(A|B)$. Show that $P_0[A|\mathcal{G}] = P[A \cap B|\mathcal{G}]/P[B|\mathcal{G}]$ on a set of P_0 -measure 1.
- (b) Suppose that \mathcal{H} is generated by a partition B_1, B_2, \dots , and let $\mathcal{G} \vee \mathcal{H} = \sigma(\mathcal{G} \cup \mathcal{H})$. Show that with probability 1,

$$P[A|\mathcal{G} \vee \mathcal{H}] = \sum_i I_{B_i} \frac{P[A \cap B_i|\mathcal{G}]}{P[B_i|\mathcal{G}]}.$$

- 34.5.** The equation (34.5) was proved by showing that the left side is a version of the right side. Prove it by showing that the right side is a version of the left side.
- 34.6.** Prove for bounded X and Y that $E[YE[X|\mathcal{G}]] = E[XE[Y|\mathcal{G}]]$.
- 34.7.** 33.9 \uparrow Generalize Theorem 34.5 by replacing X with a random vector.
- 34.8.** Assume that X is nonnegative but not necessarily integrable. Show that it is still possible to define a nonnegative random variable $E[X|\mathcal{G}]$, measurable \mathcal{G} , such that (34.1) holds. Prove versions of the monotone convergence theorem and Fatou's lemma.
- 34.9.** (a) Show for nonnegative X that $E[X|\mathcal{G}] = \int_0^\infty P[X > t|\mathcal{G}] dt$ with probability 1.
 (b) Generalize Markov's inequality: $P[|X| \geq \alpha|\mathcal{G}] \leq \alpha^{-k} E[|X|^k|\mathcal{G}]$ with probability 1.
 (c) Similarly generalize Chebyshev's and Hölder's inequalities.
- 34.10.** (a) Show that, if $\mathcal{G}_1 \subset \mathcal{G}_2$ and $E[X^2] < \infty$, then $E[(X - E[X|\mathcal{G}_2])^2] \leq E[(X - E[X|\mathcal{G}_1])^2]$. The dispersion of X about its conditional mean becomes smaller as the σ -field grows.
 (b) Define $\text{Var}[X|\mathcal{G}] = E[(X - E[X|\mathcal{G}])^2|\mathcal{G}]$. Prove that $\text{Var}[X] = E[\text{Var}[X|\mathcal{G}]] + \text{Var}[E[X|\mathcal{G}]]$.
- 34.11.** Let $\mathcal{G}_1, \mathcal{G}_2, \mathcal{G}_3$ be σ -fields in \mathcal{F} , let \mathcal{G}_{ij} be the σ -field generated by $\mathcal{G}_i \cup \mathcal{G}_j$, and let A_i be the generic set in \mathcal{G}_i . Consider three conditions:
 (i) $P[A_3|\mathcal{G}_{12}] = P[A_3|\mathcal{G}_2]$ for all A_3 .
 (ii) $P[A_1 \cap A_3|\mathcal{G}_2] = P[A_1|\mathcal{G}_2]P[A_3|\mathcal{G}_2]$ for all A_1 and A_3 .
 (iii) $P[A_1|\mathcal{G}_{23}] = P[A_1|\mathcal{G}_2]$ for all A_1 .
 If $\mathcal{G}_1, \mathcal{G}_2$, and \mathcal{G}_3 are interpreted as descriptions of the past, present, and future, respectively, (i) is a general version of the Markov property: the conditional probability of a future event A_3 given the past and present \mathcal{G}_{12} is the same as the conditional probability given the present \mathcal{G}_2 alone. Condition (iii) is the same with time reversed. And (ii) says that past and future events A_1 and A_3 are conditionally independent given the present \mathcal{G}_2 . Prove the three conditions equivalent.
- 34.12.** 33.7 34.11 \uparrow Use Example 33.10 to calculate $P[N_s = k|N_u, u \geq t]$ ($s \leq t$) for the Poisson process.
- 34.13.** Let L^2 be the Hilbert space of square-integrable random variables on (Ω, \mathcal{F}, P) . For \mathcal{G} a σ -field in \mathcal{F} , let $M_{\mathcal{G}}$ be the subspace of elements of L^2 that are measurable \mathcal{G} . Show that the operator $P_{\mathcal{G}}$ defined for $X \in L^2$ by $P_{\mathcal{G}}X = E[X|\mathcal{G}]$ is the perpendicular projection on $M_{\mathcal{G}}$.
- 34.14.** \uparrow Suppose in Problem 34.13 that $\mathcal{G} = \sigma(Z)$ for a random variable Z in L^2 . Let S_Z be the one-dimensional subspace spanned by Z . Show that S_Z may be much smaller than $M_{\sigma(Z)}$, so that $E[X|Z]$ (for $X \in L^2$) is by no means the projection of X on Z . *Hint:* Take Z the identity function on the unit interval with Lebesgue measure.

34.15. \uparrow Problem 34.13 can be turned around to give an alternative approach to conditional probability and expected value. For a σ -field \mathcal{G} in \mathcal{F} , let $P_{\mathcal{G}}$ be the perpendicular projection on the subspace $M_{\mathcal{G}}$. Show that $P_{\mathcal{G}}X$ has for $X \in L^2$ the two properties required of $E[X|\mathcal{G}]$. Use this to *define* $E[X|\mathcal{G}]$ for $X \in L^2$ and then extend it to all integrable X via approximation by random variables in L^2 . Now define conditional probability.

34.16. *Mixing sequences.* A sequence A_1, A_2, \dots of \mathcal{F} -sets in a probability space (Ω, \mathcal{F}, P) is *mixing* with constant α if

$$(34.17) \quad \lim_n P(A_n \cap E) = \alpha P(E)$$

for every E in \mathcal{F} . Then $\alpha = \lim_n P(A_n)$.

(a) Show that $\{A_n\}$ is mixing with constant α if and only if

$$(34.18) \quad \lim_n \int_{A_n} X dP = \alpha \int X dP$$

for each integrable X (measurable \mathcal{F}).

(b) Suppose that (34.17) holds for $E \in \mathcal{P}$, where \mathcal{P} is a π -system, $\Omega \in \mathcal{P}$, and $A_n \in \sigma(\mathcal{P})$ for all n . Show that $\{A_n\}$ is mixing. *Hint:* First check (34.18) for X measurable $\sigma(\mathcal{P})$ and then use conditional expected values with respect to $\sigma(\mathcal{P})$.

(c) Show that, if P_0 is a probability measure on (Ω, \mathcal{F}) and $P_0 \ll P$, then mixing is preserved if P is replaced by P_0 .

34.17. \uparrow *Application of mixing to the central limit theorem.* Let X_1, X_2, \dots be random variables on (Ω, \mathcal{F}, P) , independent and identically distributed with mean 0 and variance σ^2 , and put $S_n = X_1 + \dots + X_n$. Then $S_n/\sigma\sqrt{n} \Rightarrow N$ by the Lindeberg–Lévy theorem. Show by the steps below that this still holds if P is replaced by any probability measure P_0 on (Ω, \mathcal{F}) that P dominates. For example, the central limit theorem applies to the sums $\sum_{k=1}^n r_k(\omega)$ of Rademacher functions if ω is chosen according to the uniform density over the unit interval, and this result shows that the same is true if ω is chosen according to an arbitrary density.

Let $Y_n = S_n/\sigma\sqrt{n}$ and $Z_n = (S_n - S_{[\log n]})/\sigma\sqrt{n}$, and take \mathcal{P} to consist of the sets of the form $[(X_1, \dots, X_k) \in H]$, $k \geq 1$, $H \in \mathcal{H}^k$. Prove successively:

- (a) $P[Y_n \leq x] \rightarrow P[N \leq x]$.
- (b) $P[|Y_n - Z_n| \geq \epsilon] \rightarrow 0$.
- (c) $P[Z_n \leq x] \rightarrow P[N \leq x]$.
- (d) $P(E \cap [Z_n \leq x]) \rightarrow P(E)P[N \leq x]$ for $E \in \mathcal{P}$.
- (e) $P(E \cap [Z_n \leq x]) \rightarrow P(E)P[N \leq x]$ for $E \in \mathcal{F}$.
- (f) $P_0[Z_n \leq x] \rightarrow P[N \leq x]$.
- (g) $P_0[|Y_n - Z_n| \geq \epsilon] \rightarrow 0$.
- (h) $P_0[Y_n \leq x] \rightarrow P[N \leq x]$.

34.18. Suppose that \mathcal{G} is a sufficient subfield for the family of probability measures P_{θ} , $\theta \in \Theta$, on (Ω, \mathcal{F}) . Suppose that for each θ and A , $p(A, \omega)$ is a version of $P_{\theta}[A|\mathcal{G}]_{\omega}$, and suppose further that for each ω , $p(\cdot, \omega)$ is a probability

measure on \mathcal{F} . Define Q_θ on \mathcal{F} by $Q_\theta(A) = \int_\Omega p(A, \omega) P_\theta(d\omega)$, and show that $Q_\theta = P_\theta$.

The idea is that an observer with the information in \mathcal{G} (but ignorant of ω itself) in principle knows the values $p(A, \omega)$ because each $p(A, \cdot)$ is measurable \mathcal{G} . If he has the appropriate randomization device, he can draw an ω' from Ω according to the probability measure $p(\cdot, \omega)$, and his ω' will have the same distribution $Q_\theta = P_\theta$ that ω has. Thus, whatever the value of the unknown θ , the observer can on the basis of the information in \mathcal{G} alone, and without knowing ω itself, construct a probabilistic replica of ω .

34.19. 34.13 \uparrow In the context of the discussion on p. 252, let $\bar{\mathcal{F}}$ be the σ -field of sets of the form $\Theta \times A$ for $A \in \mathcal{F}$. Show that under the probability measure Q , \bar{i}_0 is the conditional expected value of \bar{g}_0 given $\bar{\mathcal{F}}$.

34.20. (a) In Example 34.4, take π to have density $e^{-\theta}$ over $\Theta = (0, \infty)$. Show by Theorem 34.7 that T is a minimal sufficient statistic (in the sense that $\sigma(T)$ is minimal).

(b) Let P_θ be the distribution for samples of size n from a normal distribution with parameter $\theta = (m, \sigma^2)$, $\sigma^2 > 0$, and let π put unit mass at $(0, 1)$. Show that the sample mean and variance form a minimal sufficient statistic.

SECTION 35. MARTINGALES

Definition

Let X_1, X_2, \dots be a sequence of random variables on a probability space (Ω, \mathcal{F}, P) , and let $\mathcal{F}_1, \mathcal{F}_2, \dots$ be a sequence of σ -fields in \mathcal{F} . The sequence $\{(X_n, \mathcal{F}_n): n = 1, 2, \dots\}$ is a *martingale* if these four conditions hold:

- (i) $\mathcal{F}_n \subset \mathcal{F}_{n+1}$;
- (ii) X_n is measurable \mathcal{F}_n ;
- (iii) $E[|X_n|] < \infty$;
- (iv) with probability 1,

$$(35.1) \quad E[X_{n+1} | \mathcal{F}_n] = X_n.$$

Alternatively, the sequence X_1, X_2, \dots is said to be a *martingale relative to the σ -fields $\mathcal{F}_1, \mathcal{F}_2, \dots$* . Condition (i) is expressed by saying the \mathcal{F}_n form a *filtration* and condition (ii) by saying the X_n are *adapted* to the filtration.

If X_n represents the fortune of a gambler after the n th play and \mathcal{F}_n represents his information about the game at that time, (35.1) says that his expected fortune after the next play is the same as his present fortune. Thus a martingale represents a fair game, and sums of independent random variables with mean 0 give one example. As will be seen below, martingales arise in very diverse connections.

The sequence X_1, X_2, \dots is defined to be a martingale if it is a martingale relative to *some* sequence $\mathcal{F}_1, \mathcal{F}_2, \dots$. In this case, the σ -fields $\mathcal{G}_n = \sigma(X_1, \dots, X_n)$ always work: Obviously, $\mathcal{G}_n \subset \mathcal{G}_{n+1}$ and X_n is measurable \mathcal{G}_n , and if (35.1) holds, then $E[X_{n+1} \|\mathcal{G}_n] = E[E[X_{n+1} \|\mathcal{F}_n] \|\mathcal{G}_n] = E[X_n \|\mathcal{G}_n] = X_n$ by (34.5). For these special σ -fields \mathcal{G}_n , (35.1) reduces to

$$(35.2) \quad E[X_{n+1} \|\mathcal{G}_n] = X_n.$$

Since $\sigma(X_1, \dots, X_n) \subset \mathcal{F}_n$ if and only if X_n is measurable \mathcal{F}_n for each n , the $\sigma(X_1, \dots, X_n)$ are the *smallest* σ -fields with respect to which the X_n are a martingale.

The essential condition is embodied in (35.1) and in its specialization (35.2). Condition (iii) is of course needed to ensure that $E[X_{n+1} \|\mathcal{F}_n]$ exists. Condition (iv) says that X_n is a version of $E[X_{n+1} \|\mathcal{F}_n]$; since X_n is measurable \mathcal{F}_n , the requirement reduces to

$$(35.3) \quad \int_A X_{n+1} dP = \int_A X_n dP, \quad A \in \mathcal{F}_n.$$

Since the \mathcal{F}_n are nested, $A \in \mathcal{F}_n$ implies that $\int_A X_n dP = \int_A X_{n+1} dP = \dots = \int_A X_{n+k} dP$. Therefore, X_n , being measurable \mathcal{F}_n , is a version of $E[X_{n+k} \|\mathcal{F}_n]$:

$$(35.4) \quad E[X_{n+k} \|\mathcal{F}_n] = X_n$$

with probability 1 for $k \geq 1$. Note that for $A = \Omega$, (35.3) gives

$$(35.5) \quad E[X_1] = E[X_2] = \dots.$$

The defining conditions for a martingale can also be given in terms of the differences

$$(35.6) \quad \Delta_n = X_n - X_{n-1}$$

($\Delta_1 = X_1$). By linearity, (35.1) is the same thing as

$$(35.7) \quad E[\Delta_{n+1} \|\mathcal{F}_n] = 0.$$

Note that, since $X_k = \Delta_1 + \dots + \Delta_k$ and $\Delta_k = X_k - X_{k-1}$, the sets X_1, \dots, X_n and $\Delta_1, \dots, \Delta_n$ generate the same σ -field:

$$(35.8) \quad \sigma(X_1, \dots, X_n) = \sigma(\Delta_1, \dots, \Delta_n).$$

Example 35.1. Let $\Delta_1, \Delta_2, \dots$ be independent, integrable random variables such that $E[\Delta_n] = 0$ for $n \geq 2$. If \mathcal{F}_n is the σ -field (35.8), then by independence $E[\Delta_{n+1} \|\mathcal{F}_n] = E[\Delta_{n+1}] = 0$. If Δ is another random variable,

independent of the Δ_n , and if \mathcal{F}_n is replaced by $\sigma(\Delta, \Delta_1, \dots, \Delta_n)$, then the $X_n = \Delta_1 + \dots + \Delta_n$ are still a martingale relative to the \mathcal{F}_n . It is natural and convenient in the theory to allow σ -fields \mathcal{F}_n larger than the minimal ones (35.8). ■

Example 35.2. Let (Ω, \mathcal{F}, P) be a probability space, let ν be a finite measure on \mathcal{F} , and let $\mathcal{F}_1, \mathcal{F}_2, \dots$ be a nondecreasing sequence of σ -fields in \mathcal{F} . Suppose that P dominates ν when both are restricted to \mathcal{F}_n —that is, suppose that $A \in \mathcal{F}_n$ and $P(A) = 0$ together imply that $\nu(A) = 0$. There is then a density or Radon–Nikodym derivative X_n of ν with respect to P when both are restricted to \mathcal{F}_n ; X_n is a function that is measurable \mathcal{F}_n and integrable with respect to P , and it satisfies

$$(35.9) \quad \int_A X_n dP = \nu(A), \quad A \in \mathcal{F}_n.$$

If $A \in \mathcal{F}_n$, then $A \in \mathcal{F}_{n+1}$ as well, so that $\int_A X_{n+1} dP = \nu(A)$; this and (35.9) give (35.3). Thus the X_n are a martingale with respect to the \mathcal{F}_n . ■

Example 35.3. For a specialization of the preceding example, let P be Lebesgue measure on the σ -field \mathcal{F} of Borel subsets of $\Omega = (0, 1]$, and let \mathcal{F}_n be the finite σ -field generated by the partition of Ω into dyadic intervals $(k2^{-n}, (k+1)2^{-n}]$, $0 \leq k < 2^n$. If $A \in \mathcal{F}_n$ and $P(A) = 0$, then A is empty. Hence P dominates every finite measure ν on \mathcal{F}_n . The Radon–Nikodym derivative is

$$(35.10) \quad X_n(\omega) = \frac{\nu(k2^{-n}, (k+1)2^{-n}]}{2^{-n}} \quad \text{if } \omega \in (k2^{-n}, (k+1)2^{-n}].$$

There is no need here to assume that P dominates ν when they are viewed as measures on all of \mathcal{F} . Suppose that ν is the distribution of $\sum_{k=1}^{\infty} Z_k 2^{-k}$ for independent Z_k assuming values 1 and 0 with probabilities p and $1-p$. This is the measure in Examples 31.1 and 31.3 (there denoted by μ), and for $p \neq \frac{1}{2}$, ν is singular with respect to Lebesgue measure P . It is nonetheless absolutely continuous with respect to P when both are restricted to \mathcal{F}_n . ■

Example 35.4. For another specialization of Example 35.2, suppose that ν is a probability measure Q on \mathcal{F} and that $\mathcal{F}_n = \sigma(Y_1, \dots, Y_n)$ for random variables Y_1, Y_2, \dots on (Ω, \mathcal{F}) . Suppose that under the measure P the distribution of the random vector (Y_1, \dots, Y_n) has density $p_n(y_1, \dots, y_n)$ with respect to n -dimensional Lebesgue measure and that under Q it has density $q_n(y_1, \dots, y_n)$. To avoid technicalities, assume that p_n is everywhere positive.

Then the Radon–Nikodym derivative for Q with respect to P on \mathcal{F}_n is

$$(35.11) \quad X_n = \frac{q_n(Y_1, \dots, Y_n)}{p_n(Y_1, \dots, Y_n)}.$$

To see this, note that the general element of \mathcal{F}_n is $[(Y_1, \dots, Y_n) \in H]$, $H \in \mathcal{R}^n$; by the change-of-variable formula,

$$\begin{aligned} \int_{[(Y_1, \dots, Y_n) \in H]} X_n dP &= \int_H \frac{q_n(y_1, \dots, y_n)}{p_n(y_1, \dots, y_n)} p_n(y_1, \dots, y_n) dy_1 \cdots dy_n \\ &= Q[(Y_1, \dots, Y_n) \in H]. \end{aligned}$$

In statistical terms, (35.11) is a likelihood ratio: p_n and q_n are rival densities, and the larger X_n is, the more strongly one prefers q_n as an explanation of the observation (Y_1, \dots, Y_n) . The analysis is carried out under the assumption that P is the measure actually governing the Y_n ; that is, X_n is a martingale under P and not in general under Q .

In the most common case the Y_n are independent and identically distributed under both P and Q : $p_n(y_1, \dots, y_n) = p(y_1) \cdots p(y_n)$ and $q_n(y_1, \dots, y_n) = q(y_1) \cdots q(y_n)$ for densities p and q on the line, where p is assumed everywhere positive for simplicity. Suppose that the measures corresponding to the densities p and q are not identical, so that $P[Y_n \in H] \neq Q[Y_n \in H]$ for some $H \in \mathcal{R}^1$. If $Z_n = I_{[Y_n \in H]}$, then by the strong law of large numbers, $n^{-1} \sum_{k=1}^n Z_k$ converges to $P[Y_1 \in H]$ on a set (in \mathcal{F}) of P -measure 1 and to $Q[Y_1 \in H]$ on a (disjoint) set of Q -measure 1. Thus P and Q are mutually singular on \mathcal{F} even though P dominates Q on \mathcal{F}_n . ■

Example 35.5. Suppose that Z is an integrable random variable on (Ω, \mathcal{F}, P) and that \mathcal{F}_n are nondecreasing σ -fields in \mathcal{F} . If

$$(35.12) \quad X_n = E[Z | \mathcal{F}_n],$$

then the first three conditions in the martingale definition are satisfied, and by (34.5), $E[X_{n+1} | \mathcal{F}_n] = E[E[Z | \mathcal{F}_{n+1}] | \mathcal{F}_n] = E[Z | \mathcal{F}_n] = X_n$. Thus X_n is a martingale relative to \mathcal{F}_n . ■

Example 35.6. Let N_{nk} , $n, k = 1, 2, \dots$, be an independent array of identically distributed random variables assuming the values $0, 1, 2, \dots$. Define Z_0, Z_1, Z_2, \dots inductively by $Z_0(\omega) = 1$ and $Z_n(\omega) = N_{n,1}(\omega) + \cdots + N_{n,Z_{n-1}(\omega)}(\omega)$; $Z_n(\omega) = 0$ if $Z_{n-1}(\omega) = 0$. If N_{nk} is thought of as the number of progeny of an organism, and if Z_{n-1} represents the size at time $n-1$ of a population of these organisms, then Z_n represents the size at time n . If the expected number of progeny is $E[N_{nk}] = m$, then $E[Z_n | Z_{n-1}] = Z_{n-1}m$, so that $X_n = Z_n/m^n$, $n = 0, 1, 2, \dots$, is a martingale. The sequence Z_0, Z_1, \dots is a *branching process*. ■

In the preceding definition and examples, n ranges over the positive integers. The definition makes sense if n ranges over $1, 2, \dots, N$; here conditions (ii) and (iii) are required for $1 \leq n \leq N$ and conditions (i) and (iv) only for $1 \leq n < N$. It is, in fact, clear that the definition makes sense if the indices range over an arbitrary ordered set. Although martingale theory with an interval of the line as the index set is of great interest and importance, here the index set will be discrete.

Submartingales

Random variables X_n are a *submartingale* relative to σ -fields \mathcal{F}_n if (i), (ii), and (iii) of the definition above hold and if this condition holds in place of (iv):

(iv') with probability 1,

$$(35.13) \quad E[X_{n+1} | \mathcal{F}_n] \geq X_n.$$

As before, the X_n are a submartingale if they are a submartingale with respect to some sequence \mathcal{F}_n , and the special sequence $\mathcal{F}_n = \sigma(X_1, \dots, X_n)$ works if any does. The requirement (35.13) is the same thing as

$$(35.14) \quad \int_A X_{n+1} dP \geq \int_A X_n dP, \quad A \in \mathcal{F}_n.$$

This extends inductively (see the argument for (35.4)), and so

$$(35.15) \quad E[X_{n+k} | \mathcal{F}_n] \geq X_n$$

for $k \geq 1$. Taking expected values in (35.15) gives

$$(35.16) \quad E[X_1] \leq E[X_2] \leq \dots.$$

Example 35.7. Suppose that the Δ_n are independent and integrable, as in Example 35.1, but assume that $E[\Delta_n]$ is for $n \geq 2$ nonnegative rather than 0. Then the partial sums $\Delta_1 + \dots + \Delta_n$ form a submartingale. ■

Example 35.8. Suppose that the X_n are a martingale relative to the \mathcal{F}_n . Then $|X_n|$ is measurable \mathcal{F}_n and integrable, and by Theorem 34.2(iv), $E[|X_{n+1}| | \mathcal{F}_n] \geq |E[X_{n+1} | \mathcal{F}_n]| = |X_n|$. Thus the $|X_n|$ are a submartingale relative to the \mathcal{F}_n . Note that even if X_1, \dots, X_n generate \mathcal{F}_n , in general $|X_1|, \dots, |X_n|$ will generate a σ -field smaller than \mathcal{F}_n . ■

Reversing the inequality in (35.13) gives the definition of a *supermartingale*. The inequalities in (35.14), (35.15), and (35.16) become reversed as well. The theory for supermartingales is of course symmetric to that of submartingales.

Gambling

Consider again the gambler whose fortune after the n th play is X_n and whose information about the game at that time is represented by the σ -field \mathcal{F}_n . If $\mathcal{F}_n = \sigma(X_1, \dots, X_n)$, he knows the sequence of his fortunes and nothing else, but \mathcal{F}_n could be larger. The martingale condition (35.1) stipulates that his expected or average fortune after the next play equals his present fortune, and so the martingale is the model for a *fair game*. Since the condition (35.13) for a submartingale stipulates that he stands to gain (or at least not lose) on the average, a submartingale represents a game *favorable* to the gambler. Similarly, a supermartingale represents a game *unfavorable* to the gambler.[†]

Examples of such games were studied in Section 7, and some of the results there have immediate generalizations. Start the martingale at $n = 0$, X_0 representing the gambler's initial fortune. The difference $\Delta_n = X_n - X_{n-1}$ represents the amount the gambler wins on the n th play,[‡] a negative win being of course a loss. Suppose instead that Δ_n represents the amount he wins if he puts up unit stakes. If instead of unit stakes he wagers the amount W_n on the n th play, $W_n \Delta_n$ represents his gain on that play. Suppose that $W_n \geq 0$, and that W_n is measurable \mathcal{F}_{n-1} to exclude prevision: Before the n th play the information available to the gambler is that in \mathcal{F}_{n-1} , and his choice of stake W_n must be based on this alone. For simplicity take W_n bounded. Then $W_n \Delta_n$ is integrable, and it is measurable \mathcal{F}_n if Δ_n is, and if X_n is a martingale, then $E[W_n \Delta_n | \mathcal{F}_{n-1}] = W_n E[\Delta_n | \mathcal{F}_{n-1}] = 0$ by (34.2). Thus

$$(35.17) \quad X_0 + W_1 \Delta_1 + \cdots + W_n \Delta_n$$

is a martingale relative to the \mathcal{F}_n . The sequence W_1, W_2, \dots represents a betting system, and transforming a fair game by a betting system preserves fairness; that is, transforming X_n into (35.17) preserves the martingale property.

The various betting systems discussed in Section 7 give rise to various martingales, and these martingales are not in general sums of independent random variables—are not in general the special martingales of Example 35.1. If W_n assumes only the values 0 and 1, the betting system is a selection system; see Section 7.

If the game is unfavorable to the gambler—that is, if X_n is a supermartingale—and if W_n is nonnegative, bounded, and measurable \mathcal{F}_{n-1} , then the same argument shows that (35.17) is again a supermartingale, is again unfavorable. Betting systems are thus of no avail in unfavorable games.

The stopping-time arguments of Section 7 also extend. Suppose that $\{X_n\}$ is a martingale relative to $\{\mathcal{F}_n\}$; it may have come from another martingale

[†]There is a reversal of terminology here: a subfair game (Section 7) is against the gambler, while a submartingale favors him.

[‡]The notation has, of course, changed. The F_n and X_n of Section 7 have become X_n and Δ_n .

via transformation by a betting system. Let τ be a random variable taking on nonnegative integers as values, and suppose that

$$(35.18) \quad [\tau = n] \in \mathcal{F}_n.$$

If τ is the time the gambler stops, $[\tau = n]$ is the event he stops just after the n th play, and (35.18) requires that his decision is to depend only on the information \mathcal{F}_n available to him at that time. His fortune at time n for this stopping rule is

$$(35.19) \quad X_n^* = \begin{cases} X_n & \text{if } n \leq \tau, \\ X_\tau & \text{if } n \geq \tau. \end{cases}$$

Here X_τ (which has value $X_{\tau(\omega)}(\omega)$ at ω) is the gambler's ultimate fortune, and it is his fortune for all times subsequent to τ .

The problem is to show that X_0^*, X_1^*, \dots is a martingale relative to $\mathcal{F}_0, \mathcal{F}_1, \dots$. First,

$$E[|X_n^*|] = \sum_{k=0}^{n-1} \int_{[\tau=k]} |X_k| dP + \int_{[\tau \geq n]} |X_n| dP \leq \sum_{k=0}^n E[|X_k|] < \infty.$$

Since $[\tau > n] = \Omega - [\tau \leq n] \in \mathcal{F}_n$,

$$[X_n^* \in H] = \bigcup_{k=0}^n [\tau = k, X_k \in H] \cup [\tau > n, X_n \in H] \in \mathcal{F}_n.$$

Moreover,

$$\int_A X_n^* dP = \int_{A \cap [\tau > n]} X_n dP + \int_{A \cap [\tau \leq n]} X_\tau dP$$

and

$$\int_A X_{n+1}^* dP = \int_{A \cap [\tau > n]} X_{n+1} dP + \int_{A \cap [\tau \leq n]} X_\tau dP.$$

Because of (35.3), the right sides here coincide if $A \in \mathcal{F}_n$; this establishes (35.3) for the sequence X_1^*, X_2^*, \dots , which is thus a martingale. The same kind of argument works for supermartingales.

Since $X_n^* = X_\tau$ for $n \geq \tau$, $X_n^* \rightarrow X_\tau$. As pointed out in Section 7, it is not always possible to integrate to the limit here. Let $X_n = a + \Delta_1 + \dots + \Delta_n$, where the Δ_n are independent and assume the values ± 1 with probability $\frac{1}{2}$ ($X_0 = a$), and let τ be the smallest n for which $\Delta_1 + \dots + \Delta_n = 1$. Then $E[X_0^*] = a$ and $X_\tau = a + 1$. On the other hand, if the X_n are uniformly bounded or uniformly integrable, it is possible to integrate to the limit: $E[X_\tau] = E[X_0]$.

Functions of Martingales

Convex functions of martingales are submartingales:

Theorem 35.1. (i) If X_1, X_2, \dots is a martingale relative to $\mathcal{F}_1, \mathcal{F}_2, \dots$, if φ is convex, and if the $\varphi(X_n)$ are integrable, then $\varphi(X_1), \varphi(X_2), \dots$ is a submartingale relative to $\mathcal{F}_1, \mathcal{F}_2$.

(ii) If X_1, X_2, \dots is a submartingale relative to $\mathcal{F}_1, \mathcal{F}_2, \dots$, if φ is nondecreasing and convex, and if the $\varphi(X_n)$ are integrable, then $\varphi(X_1), \varphi(X_2), \dots$ is a submartingale relative to $\mathcal{F}_1, \mathcal{F}_2, \dots$.

PROOF. In the submartingale case, $X_n \leq E[X_{n+1} | \mathcal{F}_n]$, and if φ is nondecreasing, then $\varphi(X_n) \leq \varphi(E[X_{n+1} | \mathcal{F}_n])$. In the martingale case, $X_n = E[X_{n+1} | \mathcal{F}_n]$, and so $\varphi(X_n) = \varphi(E[X_{n+1} | \mathcal{F}_n])$. If φ is convex, then by Jensen's inequality (34.7) for conditional expectations, it follows that $\varphi(E[X_{n+1} | \mathcal{F}_n]) \leq E[\varphi(X_{n+1}) | \mathcal{F}_n]$. ■

Example 35.8 is the case of part (i) for $\varphi(x) = |x|$.

Stopping Times

Let τ be a random variable taking as values positive integers or the special value ∞ . It is a *stopping time* with respect to $\{\mathcal{F}_n\}$ if $[\tau = k] \in \mathcal{F}_k$ for each finite k (see (35.18)), or, equivalently, if $[\tau \leq k] \in \mathcal{F}_k$ for each finite k . Define

$$(35.20) \quad \mathcal{F}_\tau = [A \in \mathcal{F} : A \cap [\tau \leq k] \in \mathcal{F}_k, 1 \leq k < \infty].$$

This is a σ -field, and the definition is unchanged if $[\tau \leq k]$ is replaced by $[\tau = k]$ on the right. Since clearly $[\tau = j] \in \mathcal{F}_\tau$ for finite j , τ is measurable \mathcal{F}_τ .

If $\tau(\omega) < \infty$ for all ω and $\mathcal{F}_n = \sigma(X_1, \dots, X_n)$, then $I_A(\omega) = I_A(\omega')$ for all A in \mathcal{F}_τ if and only if $X_i(\omega) = X_i(\omega')$ for $i \leq \tau(\omega) = \tau(\omega')$: The information in \mathcal{F}_τ consists of the values $\tau(\omega), X_1(\omega), \dots, X_{\tau(\omega)}(\omega)$.

Suppose now that τ_1 and τ_2 are two stopping times and $\tau_1 \leq \tau_2$. If $A \in \mathcal{F}_{\tau_1}$, then $A \cap [\tau_1 \leq k] \in \mathcal{F}_k$ and hence $A \cap [\tau_2 \leq k] = A \cap [\tau_1 \leq k] \cap [\tau_2 \leq k] \in \mathcal{F}_k$: $\mathcal{F}_{\tau_1} \subset \mathcal{F}_{\tau_2}$.

Theorem 35.2. If X_1, \dots, X_n is a submartingale with respect to $\mathcal{F}_1, \dots, \mathcal{F}_n$ and τ_1, τ_2 are stopping times satisfying $1 \leq \tau_1 \leq \tau_2 \leq n$, then X_{τ_1}, X_{τ_2} is a submartingale with respect to $\mathcal{F}_{\tau_1}, \mathcal{F}_{\tau_2}$.

This is the *optional sampling theorem*. The proof will show that X_{τ_1}, X_{τ_2} is a martingale if X_1, \dots, X_n is.

PROOF. Since the X_{τ_i} are dominated by $\sum_{k=1}^n |X_k|$, they are integrable. It is required to show that $E[X_{\tau_2} | \mathcal{F}_{\tau_1}] \geq X_{\tau_1}$, or

$$(35.21) \quad \int_A (X_{\tau_2} - X_{\tau_1}) dP \geq 0, \quad A \in \mathcal{F}_{\tau_1}.$$

But $A \in \mathcal{F}_{\tau_1}$ implies that $A \cap [\tau_1 < k \leq \tau_2] = (A \cap [\tau_1 \leq k-1]) \cap [\tau_2 \leq k-1]^c$ lies in \mathcal{F}_{k-1} . If $\Delta_k = X_k - X_{k-1}$, then

$$\begin{aligned} \int_A (X_{\tau_2} - X_{\tau_1}) dP &= \int_A \sum_{k=1}^n I_{[\tau_1 < k \leq \tau_2]} \Delta_k dP \\ &= \sum_{k=1}^n \int_{A \cap [\tau_1 < k \leq \tau_2]} \Delta_k dP \geq 0 \end{aligned}$$

by the submartingale property. ■

Inequalities

There are two inequalities that are fundamental to the theory of martingales.

Theorem 35.3. *If X_1, \dots, X_n is a submartingale, then for $\alpha > 0$,*

$$(35.22) \quad P\left[\max_{i \leq n} X_i \geq \alpha\right] \leq \frac{1}{\alpha} E[|X_n|].$$

This extends Kolmogorov's inequality: If S_1, S_2, \dots are partial sums of independent random variables with mean 0, they form a martingale; if the variances are finite, then S_1^2, S_2^2, \dots is a submartingale by Theorem 35.1(i), and (35.22) for this submartingale is exactly Kolmogorov's inequality (22.9).

PROOF. Let $\tau_2 = n$; let τ_1 be the smallest k such that $X_k \geq \alpha$, if there is one, and n otherwise. If $M_k = \max_{i \leq k} X_i$, then $[M_n \geq \alpha] \cap [\tau_1 \leq k] = [M_k \geq \alpha] \in \mathcal{F}_k$, and hence $[M_n \geq \alpha]$ is in \mathcal{F}_{τ_1} . By Theorem 35.2,

$$\begin{aligned} (35.23) \quad \alpha P[M_n \geq \alpha] &\leq \int_{[M_n \geq \alpha]} X_{\tau_1} dP \leq \int_{[M_n \geq \alpha]} X_n dP \\ &\leq \int_{[M_n \geq \alpha]} X_n^+ dP \leq E[X_n^+] \leq E[|X_n|]. \end{aligned} \quad \blacksquare$$

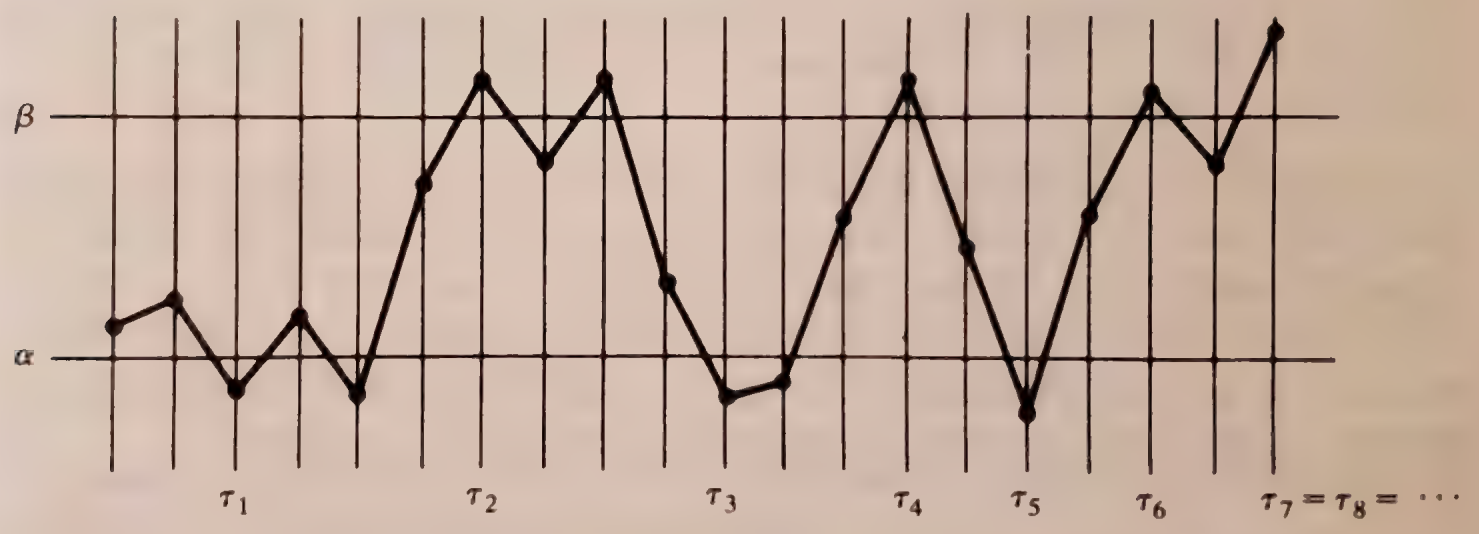
This can also be proved by imitating the argument for Kolmogorov's inequality in Section 23. For improvements to (35.22), use the other integrals

in (35.23). If X_1, \dots, X_n is a martingale, $|X_1|, \dots, |X_n|$ is a submartingale, and so (35.22) gives $P[\max_{i \leq n} |X_i| \geq \alpha] \leq \alpha^{-1} E[|X_n|]$.

The second fundamental inequality requires the notion of an *upcrossing*. Let $[\alpha, \beta]$ be an interval ($\alpha < \beta$) and let X_1, \dots, X_n be random variables. Inductively define variables τ_1, τ_2, \dots :

- τ_1 is the smallest j such that $1 \leq j \leq n$ and $X_j \leq \alpha$, and is n if there is no such j ;
- τ_k for *even* k is the smallest j such that $\tau_{k-1} < j \leq n$ and $X_j \geq \beta$, and is n if there is no such j ;
- τ_k for *odd* k exceeding 1 is the smallest j such that $\tau_{k-1} < j \leq n$ and $X_j \leq \alpha$, and is n if there is no such j .

The number U of upcrossings of $[\alpha, \beta]$ by X_1, \dots, X_n is the largest i such that $X_{\tau_{2i-1}} \leq \alpha < \beta \leq X_{\tau_{2i}}$. In the diagram, $n = 20$ and there are three upcrossings.



Theorem 35.4. For a submartingale X_1, \dots, X_n , the number U of upcrossings of $[\alpha, \beta]$ satisfies

(35.24)

$$E[U] \leq \frac{E[|X_n|] + |\alpha|}{\beta - \alpha}.$$

PROOF. Let $Y_k = \max\{0, X_k - \alpha\}$ and $\theta = \beta - \alpha$. By Theorem 35.1(ii), Y_1, \dots, Y_n is a submartingale. The τ_k are unchanged if in the definitions $X_j \leq \alpha$ is replaced by $Y_j = 0$ and $X_j \geq \beta$ by $Y_j \geq \theta$, and so U is also the number of upcrossings of $[0, \theta]$ by Y_1, \dots, Y_n . If k is even and τ_{k-1} is a stopping time, then for $j < n$,

$$[\tau_k = j] = \bigcup_{i=1}^{j-1} [\tau_{k-1} = i, Y_{i+1} < \theta, \dots, Y_{j-1} < \theta, Y_j \geq \theta]$$

lies in \mathcal{F}_j and $[\tau_k = n] = [\tau_k \leq n - 1]^c$ lies in \mathcal{F}_n , and so τ_k is also a stopping time. With a similar argument for odd k , this shows that the τ_k are all stopping times. Since the τ_k are strictly increasing until they reach n , $\tau_n = n$. Therefore,

$$Y_n = Y_{\tau_n} \geq Y_{\tau_n} - Y_{\tau_1} = \sum_{k=2}^n (Y_{\tau_k} - Y_{\tau_{k-1}}) = \Sigma_e + \Sigma_o,$$

where Σ_e and Σ_o are the sums over the even k and the odd k in the range $2 \leq k \leq n$. By Theorem 35.2, Σ_o has nonnegative expected value, and therefore, $E[Y_n] \geq E[\Sigma_e]$.

If $Y_{\tau_{2i-1}} = 0 < \theta \leq Y_{\tau_{2i}}$ (which is the same thing as $X_{\tau_{2i-1}} \leq \alpha < \beta \leq X_{\tau_{2i}}$), then the difference $Y_{\tau_{2i}} - Y_{\tau_{2i-1}}$ appears in the sum Σ_e and is at least θ . Since there are U of these differences, $\Sigma_e \geq \theta U$, and therefore $E[Y_n] \geq \theta E[U]$. In terms of the original variables, this is

$$(\beta - \alpha)E[U] \leq \int_{[X_n > \alpha]} (X_n - \alpha) dP \leq E[|X_n|] + |\alpha|. \quad \blacksquare$$

In a sense, an upcrossing of $[\alpha, \beta]$ is easy: since the X_k form a submartingale, they tend to increase. But before another upcrossing can occur, the sequence must make its way back down below α , which it resists. Think of the extreme case where the X_k are strictly increasing constants. This is reflected in the proof. Each of Σ_e and Σ_o has nonnegative expected value, but for Σ_e the proof uses the stronger inequality $E[\Sigma_e] \geq E[\theta U]$.

Convergence Theorems

The martingale convergence theorem, due to Doob, has a number of forms. The simplest one is this:

Theorem 35.5. *Let X_1, X_2, \dots be a submartingale. If $K = \sup_n E[|X_n|] < \infty$, then $X_n \rightarrow X$ with probability 1, where X is a random variable satisfying $E[|X|] \leq K$.*

PROOF. Fix α and β for the moment, and let U_n be the number of upcrossings of $[\alpha, \beta]$ by X_1, \dots, X_n . By the upcrossing theorem, $E[U_n] \leq (E[|X_n|] + |\alpha|)/(\beta - \alpha) \leq (K + |\alpha|)/(\beta - \alpha)$. Since U_n is nondecreasing and $E[U_n]$ is bounded, it follows by the monotone convergence theorem that $\sup_n U_n$ is integrable and hence finite-valued almost everywhere.

Let X^* and X_* be the limits superior and inferior of the sequence X_1, X_2, \dots ; they may be infinite. If $X_* < \alpha < \beta < X^*$, then U_n must go to infinity. Since U_n is bounded with probability 1, $P[X_* < \alpha < \beta < X^*] = 0$.

Now

$$(35.25) \quad [X_* < X^*] = \bigcup [X_* < \alpha < \beta < X^*],$$

where the union extends over all pairs of rationals α and β . The set on the left therefore has probability 0.

Thus X^* and X_* are equal with probability 1, and X_n converges to their common value X , which may be $\pm\infty$. By Fatou's lemma, $E[|X|] \leq \liminf_n E[|X_n|] \leq K$. Since it is integrable, X is finite with probability 1. ■

If the X_n form a martingale, then by (35.16) applied to the submartingale $|X_1|, |X_2|, \dots$ the $E[|X_n|]$ are nondecreasing, so that $K = \lim_n E[|X_n|]$. The hypothesis in the theorem that K be finite is essential: If $X_n = \Delta_1 + \dots + \Delta_n$, where the Δ_n are independent and assume values ± 1 with probability $\frac{1}{2}$, then X_n does not converge; $E[|X_n|]$ goes to infinity in this case.

If the X_n form a *nonnegative* martingale, then $E[|X_n|] = E[X_n] = E[X_1]$ by (35.5), and K is necessarily finite.

Example 35.9. The X_n in Example 35.6 are nonnegative, and so $X_n = Z_n/m^n \rightarrow X$, where X is nonnegative and integrable. If $m < 1$, then, since Z_n is an integer, $Z_n = 0$ for large n , and the population dies out. In this case, $X = 0$ with probability 1. Since $E[X_n] = E[X_0] = 1$, this shows that $E[X_n] \rightarrow E[X]$ may fail in Theorem 35.5. ■

Theorem 35.5 has an important application to the martingale of Example 35.5, and this requires a lemma.

Lemma. *If Z is integrable and \mathcal{F}_n are arbitrary σ -fields, then the random variables $E[Z|\mathcal{F}_n]$ are uniformly integrable.*

For the definition of uniform integrability, see (16.21). The \mathcal{F}_n must, of course, lie in the σ -field \mathcal{F} , but they need not, for example, be nondecreasing.

PROOF OF THE LEMMA. Since $|E[Z|\mathcal{F}_n]| \leq E[|Z||\mathcal{F}_n]$, Z may be assumed nonnegative. Let $A_{\alpha n} = [E[Z|\mathcal{F}_n] \geq \alpha]$. Since $A_{\alpha n} \in \mathcal{F}_n$

$$\int_{A_{\alpha n}} E[Z|\mathcal{F}_n] dP = \int_{A_{\alpha n}} Z dP.$$

It is therefore enough to find, for given ϵ , an α such that this last integral is less than ϵ for all n . Now $\int_A Z dP$ is, as a function of A , a finite measure dominated by P ; by the ϵ - δ version of absolute continuity (see (32.4)) there is a δ such that $P(A) < \delta$ implies that $\int_A Z dP < \epsilon$. But $P[E[Z|\mathcal{F}_n] \geq \alpha] \leq \alpha^{-1} E[E[Z|\mathcal{F}_n]] = \alpha^{-1} E[Z] < \delta$ for large enough α . ■

Suppose that \mathcal{F}_n are σ -fields satisfying $\mathcal{F}_1 \subset \mathcal{F}_2 \subset \cdots$. If the union $\bigcup_{n=1}^{\infty} \mathcal{F}_n$ generates the σ -field \mathcal{F}_{∞} , this is expressed by $\mathcal{F}_n \uparrow \mathcal{F}_{\infty}$. The requirement is not that \mathcal{F}_{∞} coincide with the union, but that it be generated by it.

Theorem 35.6. *If $\mathcal{F}_n \uparrow \mathcal{F}_{\infty}$ and Z is integrable, then*

$$(35.26) \quad E[Z|\mathcal{F}_n] \rightarrow E[Z|\mathcal{F}_{\infty}]$$

with probability 1.

PROOF. According to Example 35.5, the random variables $X_n = E[Z|\mathcal{F}_n]$ form a martingale relative to the \mathcal{F}_n . By the lemma, the X_n are uniformly integrable. Since $E[|X_n|] \leq E[|Z|]$, by Theorem 35.5 the X_n converge to an integrable X . The problem is to identify X with $E[Z|\mathcal{F}_{\infty}]$.

Because of the uniform integrability, it is possible (Theorem 16.14) to integrate to the limit: $\int_A X dP = \lim_n \int_A X_n dP$. If $A \in \mathcal{F}_k$ and $n \geq k$, then $\int_A X_n dP = \int_A E[Z|\mathcal{F}_n] dP = \int_A Z dP$. Therefore, $\int_A X dP = \int_A Z dP$ for all A in the π -system $\bigcup_{k=1}^{\infty} \mathcal{F}_k$; since X is measurable \mathcal{F}_{∞} , it follows by Theorem 34.1 that X is a version of $E[Z|\mathcal{F}_{\infty}]$. ■

Applications: Derivatives

Theorem 35.7. *Suppose that (Ω, \mathcal{F}, P) is a probability space, ν is a finite measure on \mathcal{F} , and $\mathcal{F}_n \uparrow \mathcal{F}_{\infty} \subset \mathcal{F}$. Suppose that P dominates ν on each \mathcal{F}_n , and let X_n be the corresponding Radon–Nikodym derivatives. Then $X_n \rightarrow X$ with probability 1, where X is integrable.*

(i) *If P dominates ν on \mathcal{F}_{∞} , then X is the corresponding Radon–Nikodym derivative.*

(ii) *If P and ν are mutually singular on \mathcal{F}_{∞} , then $X = 0$ with probability 1.*

PROOF. The situation is that of Example 35.2. The density X_n is measurable \mathcal{F}_n and satisfies (35.9). Since X_n is nonnegative, $E[|X_n|] = E[X_n] = \nu(\Omega)$, and it follows by Theorem 35.5 that X_n converges to an integrable X . The limit X is measurable \mathcal{F}_{∞} .

Suppose that P dominates ν on \mathcal{F}_{∞} and let Z be the Radon–Nikodym derivative: Z is measurable \mathcal{F}_{∞} , and $\int_A Z dP = \nu(A)$ for $A \in \mathcal{F}_{\infty}$. It follows that $\int_A Z dP = \int_A X_n dP$ for A in \mathcal{F}_n , and so $X_n = E[Z|\mathcal{F}_n]$. Now Theorem 35.6 implies that $X_n \rightarrow E[Z|\mathcal{F}_{\infty}] = Z$.

Suppose, on the other hand, that P and ν are mutually singular on \mathcal{F}_{∞} , so that there exists a set S in \mathcal{F}_{∞} such that $\nu(S) = 0$ and $P(S) = 1$. By Fatou's lemma $\int_A X dP \leq \liminf_n \int_A X_n dP$. If $A \in \mathcal{F}_k$, then $\int_A X_n dP = \nu(A)$ for $n \geq k$, and so $\int_A X dP \leq \nu(A)$ for A in the field $\bigcup_{k=1}^{\infty} \mathcal{F}_k$. It follows by the monotone class theorem that this holds for all A in \mathcal{F}_{∞} . Therefore, $\int X dP = \int_S X dP \leq \nu(S) = 0$, and X vanishes with probability 1. ■

Example 35.10. As in Example 35.3, let ν be a finite measure on the unit interval with Lebesgue measure (Ω, \mathcal{F}, P) . For \mathcal{F}_n the σ -field generated by the dyadic intervals of rank n , (35.10) gives X_n . In this case $\mathcal{F}_n \uparrow \mathcal{F}_\infty = \mathcal{F}$. For each ω and n choose the dyadic rationals $a_n(\omega) = k2^{-n}$ and $b_n(\omega) = (k+1)2^{-n}$ for which $a_n(\omega) < \omega \leq b_n(\omega)$. By Theorem 35.7, if F is the distribution function for ν , then

$$(35.27) \quad \frac{F(b_n(\omega)) - F(a_n(\omega))}{b_n(\omega) - a_n(\omega)} \rightarrow X(\omega)$$

except on a set of Lebesgue measure 0.

According to Theorem 31.2, F has a derivative F' except on a set of Lebesgue measure 0, and since the intervals $(a_n(\omega), b_n(\omega)]$ contract to ω , the difference ratio (35.27) converges almost everywhere to $F'(\omega)$ (see (31.8)). This identifies X . Since (35.27) involves intervals $(a_n(\omega), b_n(\omega)]$ of a special kind, it does not quite imply Theorem 31.2.

By Theorem 35.7, $X = F'$ is the density for ν in the absolutely continuous case, and $X = F' = 0$ (except on a set of Lebesgue measure 0) in the singular case, facts proved in a different way in Section 31. The singular case gives another example where $E[X_n] \rightarrow E[X]$ fails in Theorem 35.5. ■

Likelihood Ratios

Return to Example 35.4: $\nu = Q$ is a probability measure, $\mathcal{F}_n = \sigma(Y_1, \dots, Y_n)$ for random variables Y_n , and the Radon–Nikodym derivative or likelihood ratio X_n has the form (35.11) for densities p_n and q_n on R^n . By Theorem 35.7 the X_n converge to some X which is integrable and measurable $\mathcal{F}_\infty = \sigma(Y_1, Y_2, \dots)$.

If the Y_n are independent under P and under Q , and if the densities are different, then P and Q are mutually singular on $\sigma(Y_1, Y_2, \dots)$, as shown in Example 35.4. In this case $X = 0$ and the likelihood ratio converges to 0 on a set of P -measure 1. The statistical relevance of this is that the smaller X_n is, the more strongly one prefers P over Q as an explanation of the observation (Y_1, \dots, Y_n) , and X_n goes to 0 with probability 1 if P is in fact the measure governing the Y_n .

It might be thought that a disingenuous experimenter could bias his results by stopping at an X_n he likes—a large value if his prejudices favor Q , a small value if they favor P . This is not so, as the following analysis shows. For this argument P must dominate Q on each $\mathcal{F}_n = \sigma(Y_1, \dots, Y_n)$, but the likelihood ratio X_n need not have any special form.

Let τ be a positive-integer-valued random variable representing the time the experimenter stops. Assume that τ is finite, and to exclude prevision, assume that it is a stopping time. The σ -field \mathcal{F}_τ defined by (35.20) represents the information available at time τ , and the problem is to show that X_τ

is the likelihood ratio (Radon–Nikodym derivative) for Q with respect to P on \mathcal{F}_τ . First, X_τ is clearly measurable \mathcal{F}_τ . Second, if $A \in \mathcal{F}_\tau$, then $A \cap [\tau = n] \in \mathcal{F}_n$, and therefore

$$\int_A X_\tau dP = \sum_{n=1}^{\infty} \int_{A \cap [\tau=n]} X_n dP = \sum_{n=1}^{\infty} Q(A \cap [\tau=n]) = Q(A),$$

as required.

Reversed Martingales

A left-infinite sequence \dots, X_{-2}, X_{-1} is a martingale relative to σ -fields $\dots, \mathcal{F}_{-2}, \mathcal{F}_{-1}$ if conditions (ii) and (iii) in the definition of martingale are satisfied for $n \leq -1$ and conditions (i) and (iv) are satisfied for $n < -1$. Such a sequence is a *reversed* or *backward* martingale.

Theorem 35.8. *For a reversed martingale, $\lim_{n \rightarrow \infty} X_{-n} = X$ exists and is integrable, and $E[X] = E[X_{-n}]$ for all n .*

PROOF. The proof is almost the same as that for Theorem 35.5. Let X^* and X_* be the limits superior and inferior of the sequence X_{-1}, X_{-2}, \dots . Again (35.25) holds. Let U_n be the number of upcrossings of $[\alpha, \beta]$ by X_{-n}, \dots, X_{-1} . By the upcrossing theorem, $E[U_n] \leq (E[|X_{-1}|] + |\alpha|)/(\beta - \alpha)$. Again $E[U_n]$ is bounded, and so $\sup_n U_n$ is finite with probability 1 and the sets in (35.25) have probability 0.

Therefore, $\lim_{n \rightarrow \infty} X_{-n} = X$ exists with probability 1. By the property (35.4) for martingales, $X_{-n} = E[X_{-1} | \mathcal{F}_{-n}]$ for $n = 1, 2, \dots$. The lemma above (p. 469) implies that the X_{-n} are uniformly integrable. Therefore, X is integrable and $E[X]$ is the limit of the $E[X_{-n}]$; these all have the same value by (35.5). ■

If \mathcal{F}_n are σ -fields satisfying $\mathcal{F}_1 \supset \mathcal{F}_2 \supset \dots$, then the intersection $\bigcap_{n=1}^{\infty} \mathcal{F}_n = \mathcal{F}_0$ is also a σ -field, and the relation is expressed by $\mathcal{F}_n \downarrow \mathcal{F}_0$.

Theorem 35.9. *If $\mathcal{F}_n \downarrow \mathcal{F}_0$ and Z is integrable, then*

$$(35.28) \quad E[Z | \mathcal{F}_n] \rightarrow E[Z | \mathcal{F}_0]$$

with probability 1.

PROOF. If $X_{-n} = E[Z | \mathcal{F}_n]$, then \dots, X_{-2}, X_{-1} is a martingale relative to $\dots, \mathcal{F}_2, \mathcal{F}_1$. By the preceding theorem, $E[Z | \mathcal{F}_n]$ converges as $n \rightarrow \infty$ to an integrable X and by the lemma, the $E[Z | \mathcal{F}_n]$ are uniformly integrable. As the limit of the $E[Z | \mathcal{F}_n]$ for $n \geq k$, X is measurable \mathcal{F}_k ; k being arbitrary, X is measurable \mathcal{F}_0 .

By uniform integrability, $A \in \mathcal{F}_0$ implies that

$$\begin{aligned} \int_A X dP &= \lim_n \int_A E[Z \|\mathcal{F}_n] dP = \lim_n \int_A E[E[Z \|\mathcal{F}_n] \|\mathcal{F}_0] dP \\ &= \lim_n \int_A E[Z \|\mathcal{F}_0] dP = \int_A E[Z \|\mathcal{F}_0] dP. \end{aligned}$$

Thus X is a version of $E[Z \|\mathcal{F}_0]$. ■

Theorems 35.6 and 35.9 are parallel. There is an essential difference between Theorems 35.5 and 35.8, however. In the latter, the martingale has a last random variable, namely X_{-1} , and so it is unnecessary in proving convergence to assume the $E[|X_n|]$ bounded. On the other hand, the proof in Theorem 35.8 that X is integrable would not work for a submartingale.

Applications: de Finetti's Theorem

Let θ, X_1, X_2, \dots be random variables such that $0 \leq \theta \leq 1$ and, conditionally on θ , the X_n are independent and assume the values 1 and 0 with probabilities θ and $1 - \theta$: for u_1, \dots, u_n a sequence of 0's and 1's,

$$(35.29) \quad P[X_1 = u_1, \dots, X_n = u_n \|\theta] = \theta^s (1 - \theta)^{n-s},$$

where $s = u_1 + \dots + u_n$.

To see that such sequences exist, let θ, Z_1, Z_2, \dots be independent random variables, where θ has an arbitrarily prescribed distribution supported by $[0, 1]$ and the Z_n are uniformly distributed over $[0, 1]$. Put $X_n = I_{[Z_n \leq \theta]}$. If, for instance, $f(x) = x(1 - x) = P[Z_1 \leq x, Z_2 > x]$, then $P[X_1 = 1, X_2 = 0 \|\theta]_\omega = f(\theta(\omega))$ by (33.13). The obvious extension establishes (35.29).

Integrate (35.29):

$$(35.30) \quad P[X_1 = u_1, \dots, X_n = u_n] = E[\theta^s (1 - \theta)^{n-s}].$$

Thus $\{X_n\}$ is a mixture of Bernoulli processes. It is clear from (35.30) that the X_k are *exchangeable* in the sense that for each n the distribution of (X_1, \dots, X_n) is invariant under permutations. According to the following theorem of de Finetti, *every* exchangeable sequence is a mixture of Bernoulli sequences.

Theorem 35.10. *If the random variables X_1, X_2, \dots are exchangeable and take values 0 and 1, then there is a random variable θ for which (35.29) and (35.30) hold.*

PROOF. Let $S_m = X_1 + \cdots + X_m$. If $t \leq m$, then

$$P[S_m = t] = \sum_{u_1 \cdots u_m} P[X_1 = u_1, \dots, X_m = u_m],$$

where the sum extends over the sequences for which $u_1 + \cdots + u_m = t$. By exchangeability, the terms on the right are all equal, and since there are $\binom{m}{t}$ of them,

$$P[X_1 = u_1, \dots, X_m = u_m | S_m = t] = \binom{m}{t}^{-1}.$$

Suppose that $s \leq n \leq m$ and $u_1 + \cdots + u_n = s \leq t \leq m$; add out the u_{n+1}, \dots, u_m that sum to $t - s$:

$$\begin{aligned} P[X_1 = u_1, \dots, X_n = u_n | S_m = t] &= \binom{m-n}{t-s} / \binom{m}{t} \\ &= \frac{(t)_s (m-t)_{n-s}}{(m)_n} = f_{n,s,m} \left(\frac{t}{m} \right), \end{aligned}$$

where

$$f_{n,s,m}(x) = \prod_{i=0}^{s-1} \left(x - \frac{i}{m} \right) \prod_{i=0}^{n-s-1} \left(1 - x - \frac{i}{m} \right) / \prod_{i=0}^{n-1} \left(1 - \frac{i}{m} \right).$$

The preceding equations still hold if further constraints $S_{m+1} = t_1, \dots, S_{m+j} = t_j$ are joined to $S_m = t$. Therefore, $P[X_1 = u_1, \dots, X_n = u_n | S_m, \dots, S_{m+j}] = f_{n,s,m}(S_m/m)$.

Let $\mathcal{S}_m = \sigma(S_m, S_{m+1}, \dots)$ and $\mathcal{S} = \bigcap_m \mathcal{S}_m$. Now fix n and u_1, \dots, u_n , and suppose that $u_1 + \cdots + u_n = s$. Let $j \rightarrow \infty$ and apply Theorem 35.6: $P[X_1 = u_1, \dots, X_n = u_n | \mathcal{S}_m] = f_{n,s,m}(S_m/m)$. Let $m \rightarrow \infty$ and apply Theorem 35.9:

$$P[X_1 = u_1, \dots, X_n = u_n | \mathcal{S}]_\omega = \lim_m f_{n,s,m} \left(\frac{S_m(\omega)}{m} \right)$$

holds for ω outside a set of probability 0.

Fix such an ω and suppose that $\{S_m(\omega)/m\}$ has two distinct limit points. Since the distance from each $S_m(\omega)/m$ to the next is less than $2/m$, it follows that the set of limit points must fill a nondegenerate interval. But $\lim_k x_{m_k} = x$ implies $\lim_k f_{n,s,m_k}(x_{m_k}) = x^s(1-x)^{n-s}$, and so it follows further that $x^s(1-x)^{n-s}$ must be constant over this interval, which is impossible. Therefore, $S_m(\omega)/m$ must converge to some limit $\theta(\omega)$. This shows that $P[X_1 = u_1, \dots, X_n = u_n | \mathcal{S}] = \theta^s(1-\theta)^{n-s}$ with probability 1. Take a conditional expectation with respect to $\sigma(\theta)$, and (35.29) follows. ■

Bayes Estimation

From the Bayes point of view in statistics, the θ in (35.29) is a parameter governed by some a priori distribution known to the statistician. For given X_1, \dots, X_n , the Bayes estimate of θ is $E[\theta \| X_1, \dots, X_n]$. The problem is to show that this estimate is consistent in the sense that

$$(35.31) \quad E[\theta \| X_1, \dots, X_n] \rightarrow \theta$$

with probability 1. By Theorem 35.6, $E[\theta \| X_1, \dots, X_n] \rightarrow E[\theta \| \mathcal{F}_\infty]$, where $\mathcal{F}_\infty = \sigma(X_1, X_2, \dots)$, and so what must be shown is that $E[\theta \| \mathcal{F}_\infty] = \theta$ with probability 1.

By an elementary argument that parallels the unconditional case, it follows from (35.29) for $S_n = X_1 + \dots + X_n$ that $E[S_n \| \theta] = n\theta$ and $E[(S_n - n\theta)^2 \| \theta] = n\theta(1 - \theta)$. Hence $E[(n^{-1}S_n - \theta)^2] \leq n^{-1}$, and by Chebyshev's inequality $n^{-1}S_n$ converges in probability to θ . Therefore (Theorem 20.5), $\lim_k n_k^{-1}S_{n_k} = \theta$ with probability 1 for some subsequence. Thus $\theta = \theta'$ with probability 1 for a θ' measurable \mathcal{F}_∞ , and $E[\theta \| \mathcal{F}_\infty] = E[\theta' \| \mathcal{F}_\infty] = \theta' = \theta$ with probability 1.

A Central Limit Theorem*

Suppose X_1, X_2, \dots is a martingale relative to $\mathcal{F}_1, \mathcal{F}_2, \dots$, and put $Y_n = X_n - X_{n-1}$ ($Y_1 = X_1$), so that

$$(35.32) \quad E[Y_n \| \mathcal{F}_{n-1}] = 0.$$

View Y_n as the gambler's gain on the n th trial in a fair game. For example, if $\Delta_1, \Delta_2, \dots$ are independent and have mean 0, $\mathcal{F}_n = \sigma(\Delta_1, \dots, \Delta_n)$, W_n is measurable \mathcal{F}_{n-1} , and $Y_n = W_n \Delta_n$, then (35.32) holds (see (35.17)). A specialization of this case shows that $X_n = \sum_{k=1}^n Y_k$ need not be approximately normally distributed for large n .

Example 35.11. Suppose that Δ_n takes the values ± 1 with probability $\frac{1}{2}$ each and $W_1 = 0$, and suppose that $W_n = 1$ for $n \geq 2$ if $\Delta_1 = 1$, while $W_n = 2$ for $n \geq 2$ if $\Delta_1 = -1$. If $S_n = \Delta_2 + \dots + \Delta_n$, then X_n is S_n or $2S_n$ according as Δ_1 is $+1$ or -1 . Since S_n/\sqrt{n} has approximately the standard normal distribution, the approximate distribution of X_n/\sqrt{n} is a mixture, with equal weights, of the centered normal distributions with standard deviations 1 and 2. ■

To understand this phenomenon, consider conditional variances. Suppose for simplicity that the Y_n are bounded, and define

$$(35.33) \quad \sigma_n^2 = E[Y_n^2 \| \mathcal{F}_{n-1}]$$

*This topic, which requires the limit theory of Chapter 5, may be omitted.

(take $\mathcal{F}_0 = \{\emptyset, \Omega\}$). Consider the stopping times

$$(35.34) \quad \nu_t = \min \left[n: \sum_{k=1}^n \sigma_k^2 \geq t \right].$$

Under appropriate conditions, X_{ν_t}/\sqrt{t} will be approximately normally distributed for large t . Consider the preceding example. Roughly: If $\Delta_1 = +1$, then $\sum_{k=1}^n \sigma_k^2 = n-1$, and so $\nu_t \approx t$ and $X_{\nu_t}/\sqrt{t} \approx S_t/\sqrt{t}$; if $\Delta_1 = -1$, then $\sum_{k=1}^n \sigma_k^2 = 4(n-1)$, and so $\nu_t \approx t/4$ and $X_{\nu_t}/\sqrt{t} \approx 2S_{t/4}/\sqrt{t} = S_{t/4}/\sqrt{t/4}$. In either case, X_{ν_t}/\sqrt{t} approximately follows the standard normal law.

If the n th play takes σ_n^2 units of time, then ν_t is essentially the number of plays that take place during the first t units of time. This change of the time scale stabilizes the rate at which money changes hands.

Theorem 35.11. *Suppose the $Y_n = X_n - X_{n-1}$ are uniformly bounded and satisfy (35.32), and assume that $\sum_n \sigma_n^2 = \infty$ with probability 1. Then $X_{\nu_t}/\sqrt{t} \Rightarrow N$.*

This will be deduced from a more general result, one that contains the Lindeberg theorem. Suppose that, for each n , X_{n1}, X_{n2}, \dots is a martingale with respect to $\mathcal{F}_{n1}, \mathcal{F}_{n2}, \dots$. Define $Y_{nk} = X_{nk} - X_{n,k-1}$, suppose the Y_{nk} have second moments, and put $\sigma_{nk}^2 = E[Y_{nk}^2 | \mathcal{F}_{n,k-1}]$ ($\mathcal{F}_{n0} = \{\emptyset, \Omega\}$). The probability space may vary with n . If the martingale is originally defined only for $1 \leq k \leq r_n$, take $Y_{nk} = 0$ and $\mathcal{F}_{nk} = \mathcal{F}_{nr_n}$ for $k > r_n$. Assume that $\sum_{k=1}^\infty Y_{nk}$ and $\sum_{k=1}^\infty \sigma_{nk}^2$ converge with probability 1.

Theorem 35.12. *Suppose that*

$$(35.35) \quad \sum_{k=1}^\infty \sigma_{nk}^2 \rightarrow_P \sigma^2,$$

where σ is a positive constant, and that

$$(36.36) \quad \sum_{k=1}^\infty E[Y_{nk}^2 I_{[|Y_{nk}| \geq \epsilon]}] \rightarrow 0$$

for each ϵ . Then $\sum_{k=1}^\infty Y_{nk} \Rightarrow \sigma N$.

PROOF OF THEOREM 35.11. The proof will be given for t going to infinity through the integers.[†] Let $Y_{nk} = I_{[\nu_n \geq k]} Y_k / \sqrt{n}$ and $\mathcal{F}_{nk} = \mathcal{F}_k$. From $[\nu_n \geq k] = [\sum_{j=1}^{k-1} \sigma_j^2 < n] \in \mathcal{F}_{k-2}$ follow $E[Y_{nk} | \mathcal{F}_{n,k-1}] = 0$ and $\sigma_{nk}^2 = E[Y_{nk}^2 | \mathcal{F}_{n,k-1}] = I_{[\nu_n \geq k]} \sigma_k^2 / n$. If K bounds the $|Y_k|$, then $1 \leq \sum_{k=1}^\infty \sigma_{nk}^2 = n^{-1} \sum_{k=1}^{\nu_n} \sigma_k^2 \leq 1 + K^2/n$, so that (35.35) holds for $\sigma = 1$. For n large enough that $K/\sqrt{n} < \epsilon$, the sum in (35.36) vanishes. Theorem 35.12 therefore applies, and $\sum_{k=1}^{\nu_n} Y_k / \sqrt{n} = \sum_{k=1}^\infty Y_{nk} \Rightarrow N$. ■

[†]For the general case, first check that the proof of Theorem 35.12 goes through without change if n is replaced by a parameter going continuously to infinity.

PROOF OF THEOREM 35.12. Assume at first that there is a constant c such that

$$(35.37) \quad \sum_{k=1}^{\infty} \sigma_{nk}^2 \leq c,$$

which in fact suffices for the application to Theorem 35.11.

Write $S_k = \sum_{j=1}^k Y_{nj}$ ($S_0 = 0$), $S_{\infty} = \sum_{j=1}^{\infty} Y_{nj}$, $\Sigma_k = \sum_{j=1}^k \sigma_{nj}^2$ ($\Sigma_0 = 0$), and $\Sigma_{\infty} = \sum_{j=1}^{\infty} \sigma_{nj}^2$; the dependence on n is suppressed in the notation. To prove $E[e^{itS_{\infty}}] \rightarrow e^{-\frac{1}{2}t^2\sigma^2}$, observe first that

$$\begin{aligned} & \left| E[e^{itS_{\infty}} - e^{-\frac{1}{2}t^2\sigma^2}] \right| \\ &= \left| E[e^{itS_{\infty}}(1 - e^{\frac{1}{2}t^2\Sigma_{\infty}}e^{-\frac{1}{2}t^2\sigma^2}) + e^{-\frac{1}{2}t^2\sigma^2}(e^{\frac{1}{2}t^2\Sigma_{\infty}}e^{itS_{\infty}} - 1)] \right| \\ &\leq E[|1 - e^{\frac{1}{2}t^2\Sigma_{\infty}}e^{-\frac{1}{2}t^2\sigma^2}|] + |E[e^{\frac{1}{2}t^2\Sigma_{\infty}}e^{itS_{\infty}} - 1]| = A + B. \end{aligned}$$

The term A on the right goes to 0 as $n \rightarrow \infty$, because by (35.35) and (35.37) the integrand is bounded and goes to 0 in probability.

The integrand in B is

$$\sum_{k=1}^{\infty} e^{itS_{k-1}}(e^{itY_{nk}} - e^{-\frac{1}{2}t^2\sigma_{nk}^2})e^{\frac{1}{2}t^2\Sigma_k},$$

because the m th partial sum here telescopes to $e^{itS_m}e^{\frac{1}{2}t^2\Sigma_m} - 1$. Since, by (35.37), this partial sum is bounded uniformly in m , and since S_{k-1} and Σ_k are measurable $\mathcal{F}_{n,k-1}$, it follows (Theorem 16.7) that

$$\begin{aligned} B &= \left| \sum_{k=1}^{\infty} E[e^{itS_{k-1}}e^{\frac{1}{2}t^2\Sigma_k}(e^{itY_{nk}} - e^{-\frac{1}{2}t^2\sigma_{nk}^2})] \right| \\ &\leq \sum_{k=1}^{\infty} \left| E[e^{itS_{k-1}}e^{\frac{1}{2}t^2\Sigma_k} E[e^{itY_{nk}} - e^{-\frac{1}{2}t^2\sigma_{nk}^2} | \mathcal{F}_{n,k-1}]] \right| \\ &\leq e^{\frac{1}{2}t^2c} \sum_{k=1}^{\infty} E[|E[e^{itY_{nk}} - e^{-\frac{1}{2}t^2\sigma_{nk}^2} | \mathcal{F}_{n,k-1}]]|. \end{aligned}$$

To complete the proof (under the temporary assumption (35.37)), it is enough to show that this last sum goes to 0.

By (26.4₂),

$$(35.38) \quad e^{itY_{nk}} = 1 + itY_{nk} - \frac{1}{2}t^2Y_{nk}^2 + \theta,$$

where (write $I_{nk} = I_{|Y_{nk}| \geq \epsilon}$) and let K_t bound t^2 and $|t|^3$)

$$|\theta| \leq \min\{|tY_{nk}|^3, |tY_{nk}|^2\} \leq K_t(Y_{nk}^2 I_{nk} + \epsilon Y_{nk}^2).$$

And

(38.39)
$$e^{-\frac{1}{2}t^2\sigma_{nk}^2} = 1 - \frac{1}{2}t^2\sigma_{nk}^2 + \theta',$$

where (use (27.15) and increase K_t)

$$|\theta'| \leq \left(\frac{1}{2}t^2\sigma_{nk}^2\right)^2 e^{\frac{1}{2}t^2\sigma_{nk}^2} \leq t^4\sigma_{nk}^4 e^{\frac{1}{2}t^2c} \leq K_t\sigma_{nk}^4.$$

Because of the condition $E[Y_{nk}|\mathcal{F}_{n,k-1}] = 0$ and the definition of σ_{nk}^2 , the right sides of (35.38) and (35.39), minus θ and θ' , respectively, have the same conditional expected value given $\mathcal{F}_{n,k-1}$. By (35.37), therefore,

$$\begin{aligned} &\sum_{k=1}^\infty E\left[\left|E\left[e^{itY_{nk}} - e^{-\frac{1}{2}t^2\sigma_{nk}^2} \middle| \mathcal{F}_{n,k-1}\right]\right|\right] \\ &\leq K_t \sum_{k=1}^\infty \left(E[Y_{nk}^2 I_{nk}] + \epsilon E[\sigma_{nk}^2] + E[\sigma_{nk}^4]\right) \\ &\leq K_t \left(\sum_{k=1}^\infty E[Y_{nk}^2 I_{nk}] + \epsilon c + cE\left[\sup_{k\geq 1} \sigma_{nk}^2\right]\right). \end{aligned}$$

Since $\sigma_{nk}^2 \leq E[\epsilon^2 + Y_{nk}^2 I_{nk}|\mathcal{F}_{n,k-1}] \leq \epsilon^2 + \sum_{j=1}^\infty E[Y_{nj}^2 I_{nj}|\mathcal{F}_{n,j-1}]$, it follows by (35.36) that the last expression above is, in the limit, at most $K_t(\epsilon c + c\epsilon^2)$. Since ϵ is arbitrary, this completes the proof of the theorem under the assumption (35.37).

To remove this assumption, take $c > \sigma^2$, define $A_{nk} = [\sum_{j=1}^k \sigma_{nj}^2 \leq c]$ and $A_{n\infty} = [\sum_{j=1}^\infty \sigma_{nj}^2 \leq c]$, and take $Z_{nk} = Y_{nk} I_{A_{nk}}$. From $A_{nk} \in \mathcal{F}_{n,k-1}$ follow $E[Z_{nk}|\mathcal{F}_{n,k-1}] = 0$ and $\tau_{nk}^2 = E[Z_{nk}^2|\mathcal{F}_{n,k-1}] = I_{A_{nk}}\sigma_{nk}^2$. Since $\sum_{j=1}^\infty \tau_{nj}^2$ is $\sum_{j=1}^k \sigma_{nj}^2$ on $A_{nk} - A_{n,k+1}$ and $\sum_{j=1}^\infty \sigma_{nj}^2$ on $A_{n\infty}$, the Z -array satisfies (35.37). Now $P(A_{n\infty}) \rightarrow 1$ by (35.35), and on $A_{n\infty}$, $\tau_{nk}^2 = \sigma_{nk}^2$ for all k , so that the Z -array satisfies (35.35). And it satisfies (35.36) because $|Z_{nk}| \leq |Y_{nk}|$. Therefore, by the case already treated, $\sum_{k=1}^\infty Z_{nk} \Rightarrow \sigma N$. But since $\sum_{k=1}^\infty Y_{nk}$ coincides with this last sum on $A_{n\infty}$, it, too, is asymptotically normal. ■

PROBLEMS

- 35.1. Suppose that $\Delta_1, \Delta_2, \dots$ are independent random variables with mean 0. Let $X_1 = \Delta_1$ and $X_{n+1} = X_n + \Delta_{n+1} f_n(X_1, \dots, X_n)$, and suppose that the X_n are integrable. Show that $\{X_n\}$ is a martingale. The martingales of gambling have this form.
- 35.2. Let Y_1, Y_2, \dots be independent random variables with mean 0 and variance σ^2 . Let $X_n = (\sum_{k=1}^n Y_k)^2 - n\sigma^2$ and show that $\{X_n\}$ is a martingale.

- 35.3.** Suppose that $\{Y_n\}$ is a finite-state Markov chain with transition matrix $[p_{ij}]$. Suppose that $\sum_j p_{ij}x(j) = \lambda x(i)$ for all i (the $x(i)$ are the components of a right eigenvector of the transition matrix). Put $X_n = \lambda^{-n}x(Y_n)$ and show that $\{X_n\}$ is a martingale.
- 35.4.** Suppose that Y_1, Y_2, \dots are independent, positive random variables and that $E[Y_n] = 1$. Put $X_n = Y_1 \cdots Y_n$.
- (a) Show that $\{X_n\}$ is a martingale and converges with probability 1 to an integrable X .
- (b) Suppose specifically that Y_n assumes the values $\frac{1}{2}$ and $\frac{3}{2}$ with probability $\frac{1}{2}$ each. Show that $X = 0$ with probability 1. This gives an example where $E[\prod_{n=1}^{\infty} Y_n] \neq \prod_{n=1}^{\infty} E[Y_n]$ for independent, integrable, positive random variables. Show, however, that $E[\prod_{n=1}^{\infty} Y_n] \leq \prod_{n=1}^{\infty} E[Y_n]$ always holds.
- 35.5.** Suppose that X_1, X_2, \dots is a martingale satisfying $E[X_1] = 0$ and $E[X_n^2] < \infty$. Show that $E[(X_{n+r} - X_n)^2] = \sum_{k=1}^r E[(X_{n+k} - X_{n+k-1})^2]$ (the variance of the sum is the sum of the variances). Assume that $\sum_n E[(X_n - X_{n-1})^2] < \infty$ and prove that X_n converges with probability 1. Do this first by Theorem 35.5 and then (see Theorem 22.6) by Theorem 35.3.
- 35.6.** Show that a submartingale X_n can be represented as $X_n = Y_n + Z_n$, where Y_n is a martingale and $0 \leq Z_1 \leq Z_2 \leq \dots$. *Hint:* Take $X_0 = 0$ and $\Delta_n = X_n - X_{n-1}$, and define $Z_n = \sum_{k=1}^n E[\Delta_k | \mathcal{F}_{k-1}]$ ($\mathcal{F}_0 = \{0, \Omega\}$).
- 35.7.** If X_1, X_2, \dots is a martingale and bounded either above or below, then $\sup_n E[|X_n|] < \infty$.
- 35.8.** \uparrow Let $X_n = \Delta_1 + \dots + \Delta_n$, where the Δ_n are independent and assume the values ± 1 with probability $\frac{1}{2}$ each. Let τ be the smallest n such that $X_n = 1$ and define X_n^* by (35.19). Show that the hypotheses of Theorem 35.5 are satisfied by $\{X_n^*\}$ but that it is impossible to integrate to the limit. *Hint:* Use (7.8) and Problem 35.7.
- 35.9.** Let X_1, X_2, \dots be a martingale, and assume that $|X_1(\omega)|$ and $|X_n(\omega) - X_{n-1}(\omega)|$ are bounded by a constant independent of ω and n . Let τ be a stopping time with finite mean. Show that X_τ is integrable and that $E[X_\tau] = E[X_1]$.
- 35.10.** 35.8 35.9 \uparrow Use the preceding result to show that the τ in Problem 35.8 has infinite mean. Thus the waiting time until a symmetric random walk moves one step up from the starting point has infinite expected value.
- 35.11.** Let X_1, X_2, \dots be a Markov chain with countable state space S and transition probabilities p_{ij} . A function φ on S is excessive or superharmonic if $\varphi(i) \geq \sum_j p_{ij}\varphi(j)$. Show by martingale theory that $\varphi(X_n)$ converges with probability 1 if φ is bounded and excessive. Deduce from this that if the chain is irreducible and persistent, then φ must be constant. Compare Problem 8.34.
- 35.12.** \uparrow A function φ on the integer lattice in R^k is superharmonic if for each lattice point x , $\varphi(x) \geq (2k)^{-1} \sum \varphi(y)$, the sum extending over the $2k$ nearest neighbors y . Show for $k = 1$ and $k = 2$ that a bounded superharmonic function is constant. Show for $k \geq 3$ that there exist nonconstant bounded harmonic functions.

35.13. 32.7 32.9 \uparrow Let (Ω, \mathcal{F}, P) be a probability space, let ν be a finite measure on \mathcal{F} , and suppose that $\mathcal{F}_n \uparrow \mathcal{F}_\infty \subset \mathcal{F}$. For $n \leq \infty$, let X_n be the Radon-Nikodym derivative with respect to P of the absolutely continuous part of ν when P and ν are both restricted to \mathcal{F}_n . The problem is to extend Theorem 35.7 by showing that $X_n \rightarrow X_\infty$ with probability 1.

(a) For $n \leq \infty$, let

$$\nu(A) = \int_A X_n dP + \sigma_n(A), \quad A \in \mathcal{F}_n,$$

be the decomposition of ν into absolutely continuous and singular parts with respect to P on \mathcal{F}_n . Show that X_1, X_2, \dots is a supermartingale and converges with probability 1.

(b) Let

$$\sigma_\infty(A) = \int_A Z_n dP + \sigma'_n(A), \quad A \in \mathcal{F}_n,$$

be the decomposition of σ_∞ into absolutely continuous and singular parts with respect to P on \mathcal{F}_n . Let $Y_n = E[X_\infty | \mathcal{F}_n]$, and prove

$$\int_A (Y_n + Z_n) dP + \sigma'_n(A) = \int_A X_n dP + \sigma_n(A), \quad A \in \mathcal{F}_n.$$

Conclude that $Y_n + Z_n = X_n$ with probability 1. Since Y_n converges to X_∞ , Z_n converges with probability 1 to some Z . Show that $\int_A Z dP \leq \sigma_\infty(A)$ for $A \in \mathcal{F}_\infty$, and conclude that $Z = 0$ with probability 1.

35.14. (a) Show that $\{X_n\}$ is a martingale with respect to $\{\mathcal{F}_n\}$ if and only if, for all n and all stopping times τ such that $\tau \leq n$, $E[X_n | \mathcal{F}_\tau] = X_\tau$.

(b) Show that, if $\{X_n\}$ is a martingale and τ is a bounded stopping time, then $E[X_\tau] = E[X_1]$.

35.15. 31.9 \uparrow Suppose that $\mathcal{F}_n \uparrow \mathcal{F}_\infty$ and $A \in \mathcal{F}_\infty$, and prove that $P[A | \mathcal{F}_n] \rightarrow I_A$ with probability 1. Compare Lebesgue's density theorem.

35.16. Theorems 35.6 and 35.9 have analogues in Hilbert space. For $n \leq \infty$, let P_n be the perpendicular projection on a subspace M_n . Then $P_n x \rightarrow P_\infty x$ for all x if either (a) $M_1 \subset M_2 \subset \dots$ and M_∞ is the closure of $\bigcup_{n < \infty} M_n$ or (b) $M_1 \supset M_2 \supset \dots$ and $M_\infty = \bigcap_{n < \infty} M_n$.

35.17. Suppose that θ has an arbitrary distribution, and suppose that, conditionally on θ , the random variables Y_1, Y_2, \dots are independent and normally distributed with mean θ and variance σ^2 . Construct such a sequence $\{\theta, Y_1, Y_2, \dots\}$. Prove (35.31).

35.18. It is shown on p. 471 that optional stopping has no effect on likelihood ratios. This is not true of tests of significance. Suppose that X_1, X_2, \dots are independent and identically distributed and assume the values 1 and 0 with probabilities p and $1 - p$. Consider the null hypothesis that $p = \frac{1}{2}$ and the alternative that $p > \frac{1}{2}$. The usual .05-level test of significance is to reject the null

hypothesis if

$$(35.40) \quad \frac{2}{\sqrt{n}} (X_1 + \cdots + X_n - \tfrac{1}{2}n) > 1.645.$$

For this test the chance of falsely rejecting the null hypothesis is approximately $P[N > 1.645] \approx .05$ if n is large and fixed. Suppose that n is not fixed in advance of sampling, and show by the law of the iterated logarithm that, even if p is, in fact, $\frac{1}{2}$, there are with probability 1 infinitely many n for which (35.40) holds.

- 35.19.** (a) Suppose that (35.32) and (35.33) hold. Suppose further that, for constants s_n^2 , $s_n^{-2} \sum_{k=1}^n \sigma_k^2 \rightarrow_P 1$ and $s_n^{-2} \sum_{k=1}^n E[Y_k^2 I_{\{|Y_k| \geq s_n\}}] \rightarrow 0$, and show that $s_n^{-1} \sum_{k=1}^n Y_k \Rightarrow N$. *Hint:* Simplify the proof of Theorem 35.11.
 (b) *The Lindeberg-Lévy theorem for martingales.* Suppose that

$$\dots, Y_{-1}, Y_0, Y_1, \dots$$

is stationary and ergodic (p. 494) and that

$$E[Y_k^2] < \infty \quad \text{and} \quad E[Y_k | Y_{k-1}, Y_{k-2}, \dots] = 0.$$

Prove that $\sum_{k=1}^n Y_k / \sqrt{n}$ is asymptotically normal. *Hint:* Use Theorem 36.4 and the remark following the statement of Lindeberg's Theorem 27.2.

- 35.20.** 24.4 \uparrow Suppose that the σ -field \mathcal{F}_∞ in Problem 24.4 is trivial. Deduce from Theorem 35.9 that $P[A | T^{-n} \mathcal{F}] \rightarrow P[A | \mathcal{F}_\infty] = P(A)$ with probability 1, and conclude that T is mixing.

CHAPTER 7

Stochastic Processes

SECTION 36. KOLMOGOROV'S EXISTENCE THEOREM

Stochastic Processes

A *stochastic process* is a collection $[X_t: t \in T]$ of random variables on a probability space (Ω, \mathcal{F}, P) . The sequence of gambler's fortunes in Section 7, the sequences of independent random variables in Section 22, the martingales in Section 35—all these are stochastic processes for which $T = \{1, 2, \dots\}$. For the Poisson process $[N_t: t \geq 0]$ of Section 23, $T = [0, \infty)$. For all these processes the points of T are thought of as representing *time*. In most cases, T is the set of integers and time is *discrete*, or else T is an interval of the line and time is *continuous*. For the general theory of this section, however, T can be quite arbitrary.

Finite-Dimensional Distributions

A process is usually described in terms of distributions it induces in Euclidean spaces. For each k -tuple (t_1, \dots, t_k) of distinct elements of T , the random vector $(X_{t_1}, \dots, X_{t_k})$ has over R^k some distribution μ_{t_1, \dots, t_k} :

$$(36.1) \quad \mu_{t_1, \dots, t_k}(H) = P[(X_{t_1}, \dots, X_{t_k}) \in H], \quad H \in \mathcal{R}^k.$$

These probability measures μ_{t_1, \dots, t_k} are the *finite-dimensional distributions* of the stochastic process $[X_t: t \in T]$. The system of finite-dimensional distributions does not completely determine the properties of the process. For example, the Poisson process $[N_t: t \geq 0]$ as defined by (23.5) has sample paths (functions $N_t(\omega)$ with ω fixed and t varying) that are step functions. But (23.28) defines a process that has the same finite-dimensional distributions and has sample paths that are *not* step functions. Nevertheless, the first step in a general theory is to construct processes for given systems of finite-dimensional distributions.

Now (36.1) implies two consistency properties of the system $\mu_{t_1 \dots t_k}$. Suppose the H in (36.1) has the form $H = H_1 \times \cdots \times H_k$ ($H_i \in \mathcal{R}^1$), and consider a permutation π of $(1, 2, \dots, k)$. Since $[(X_{t_1}, \dots, X_{t_k}) \in (H_1 \times \cdots \times H_k)]$ and $[(X_{t_{\pi 1}}, \dots, X_{t_{\pi k}}) \in (H_{\pi 1} \times \cdots \times H_{\pi k})]$ are the same event, it follows by (36.1) that

$$(36.2) \quad \mu_{t_1 \dots t_k}(H_1 \times \cdots \times H_k) = \mu_{t_{\pi 1} \dots t_{\pi k}}(H_{\pi 1} \times \cdots \times H_{\pi k}).$$

For example, if $\mu_{s,t} = \nu \times \nu'$, then necessarily $\mu_{t,s} = \nu' \times \nu$.

The second consistency condition is

$$(36.3) \quad \mu_{t_1 \dots t_{k-1}}(H_1 \times \cdots \times H_{k-1}) = \mu_{t_1 \dots t_{k-1} t_k}(H_1 \times \cdots \times H_{k-1} \times R^1).$$

This is clear because $(X_{t_1}, \dots, X_{t_{k-1}})$ lies in $H_1 \times \cdots \times H_{k-1}$ if and only if $(X_{t_1}, \dots, X_{t_{k-1}}, X_{t_k})$ lies in $H_1 \times \cdots \times H_{k-1} \times R^1$.

Measures $\mu_{t_1 \dots t_k}$ coming from a process $[X_t: t \in T]$ via (36.1) necessarily satisfy (36.2) and (36.3). *Kolmogorov's existence theorem* says conversely that if a given system of measures satisfies the two consistency conditions, then there exists a stochastic process having these finite-dimensional distributions. The proof is a construction, one which is more easily understood if (36.2) and (36.3) are combined into a single condition.

Define $\varphi_\pi: R^k \rightarrow R^k$ by

$$\varphi_\pi(x_1, \dots, x_k) = (x_{\pi^{-1}1}, \dots, x_{\pi^{-1}k});$$

φ_π applies the permutation π to the coordinates (for example, if π sends x_3 to first position, then $\pi^{-1}1 = 3$). Since $\varphi_\pi^{-1}(H_1 \times \cdots \times H_k) = H_{\pi 1} \times \cdots \times H_{\pi k}$, it follows from (36.2) that

$$\mu_{t_{\pi 1} \dots t_{\pi k}} \varphi_\pi^{-1}(H) = \mu_{t_1 \dots t_k}(H)$$

for rectangles H . But then

$$(36.4) \quad \mu_{t_1 \dots t_k} = \mu_{t_{\pi 1} \dots t_{\pi k}} \varphi_\pi^{-1}.$$

Similarly, if $\varphi: R^k \rightarrow R^{k-1}$ is the projection $\varphi(x_1, \dots, x_k) = (x_1, \dots, x_{k-1})$, then (36.3) is the same thing as

$$(36.5) \quad \mu_{t_1 \dots t_{k-1}} = \mu_{t_1 \dots t_k} \varphi^{-1}.$$

The conditions (36.4) and (36.5) have a common extension. Suppose that (u_1, \dots, u_m) is an m -tuple of distinct elements of T and that each element of (t_1, \dots, t_k) is also an element of (u_1, \dots, u_m) . Then (t_1, \dots, t_k) must be the initial segment of some permutation of (u_1, \dots, u_m) ; that is, $k \leq m$ and there

is a permutation π of $(1, 2, \dots, m)$ such that

$$(u_{\pi^{-1}1}, \dots, u_{\pi^{-1}m}) = (t_1, \dots, t_k, t_{k+1}, \dots, t_m),$$

where t_{k+1}, \dots, t_m are elements of (u_1, \dots, u_m) that do not appear in (t_1, \dots, t_k) . Define $\psi: R^m \rightarrow R^k$ by

$$(36.6) \quad \psi(x_1, \dots, x_m) = (x_{\pi^{-1}1}, \dots, x_{\pi^{-1}k});$$

ψ applies π to the coordinates and then projects onto the first k of them. Since $\psi(X_{u_1}, \dots, X_{u_m}) = (X_{t_1}, \dots, X_{t_k})$,

$$(36.7) \quad \mu_{t_1 \dots t_k} = \mu_{u_1 \dots u_m} \psi^{-1}.$$

This contains (36.4) and (36.5) as special cases, but as ψ is a coordinate permutation followed by a sequence of projections of the form $(x_1, \dots, x_l) \rightarrow (x_1, \dots, x_{l-1})$, it is also a consequence of these special cases.

Product Spaces

The standard construction of the general process involves product spaces. Let T be an arbitrary index set, and let R^T be the collection of all real functions on T —all maps from T into the real line. If $T = \{1, 2, \dots, k\}$, a real function on T can be identified with a k -tuple (x_1, \dots, x_k) of real numbers, and so R^T can be identified with k -dimensional Euclidean space R^k . If $T = \{1, 2, \dots\}$, a real function on T is a sequence $\{x_1, x_2, \dots\}$ of real numbers. If T is an interval, R^T consists of all real functions, however irregular, on the interval. The theory of R^T is an elaboration of the theory of the analogous but simpler space S^∞ of Section 2 (p. 27).

Whatever the set T may be, an element of R^T will be denoted x . The value of x at t will be denoted $x(t)$ or x_t , depending on whether x is viewed as a function of t with domain T or as a vector with components indexed by the elements t of T . Just as R^k can be regarded as the Cartesian product of k copies of the real line, R^T can be regarded as a *product space*—a product of copies of the real line, one copy for each t in T .

For each t define a mapping $Z_t: R^T \rightarrow R^1$ by

$$(36.8) \quad Z_t(x) = x(t) = x_t.$$

The Z_t are called the *coordinate functions* or *projections*. When later on a probability measure has been defined on R^T , the Z_t will be random variables, the *coordinate variables*. Frequently, the value $Z_t(x)$ is instead denoted $Z(t, x)$. If x is fixed, $Z(\cdot, x)$ is a real function on T and is, in fact, nothing other than $x(\cdot)$ —that is, x itself. If t is fixed, $Z(t, \cdot)$ is a real function on R^T and is identical with the function Z_t defined by (36.8).

There is a natural generalization to R^T of the idea of the σ -field of k -dimensional Borel sets. Let \mathcal{R}^T be the σ -field generated by all the coordinate functions Z_t , $t \in T$: $\mathcal{R}^T = \sigma[Z_t: t \in T]$. It is generated by the sets of the form

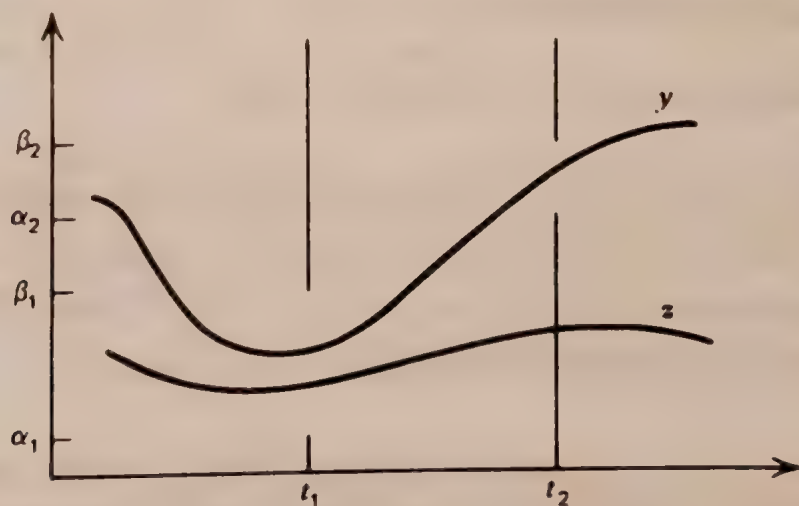
$$[x \in R^T: Z_t(x) \in H] = [x \in R^T: x_t \in H]$$

for $t \in T$ and $H \in \mathcal{R}^1$. If $T = \{1, 2, \dots, k\}$, then \mathcal{R}^T coincides with \mathcal{R}^k .

Consider the class \mathcal{R}_0^T consisting of the sets of the form

$$(36.9) \quad \begin{aligned} A &= [x \in R^T: (Z_{t_1}(x), \dots, Z_{t_k}(x)) \in H] \\ &= [x \in R^T: (x_{t_1}, \dots, x_{t_k}) \in H], \end{aligned}$$

where k is an integer, (t_1, \dots, t_k) is a k -tuple of distinct points of T , and $H \in \mathcal{R}^k$. Sets of this form, elements of \mathcal{R}_0^T , are called *finite-dimensional sets*, or *cylinders*. Of course, \mathcal{R}_0^T generates \mathcal{R}^T . Now \mathcal{R}_0^T is not a σ -field, does not coincide with \mathcal{R}^T (unless T is finite), but the following argument shows that it is a field.



If T is an interval, the cylinder $[x \in R^T: \alpha_1 < x(t_1) \leq \beta_1, \alpha_2 < x(t_2) \leq \beta_2]$ consists of the functions that go through the two gates shown; y lies in the cylinder and z does not (they need not be continuous functions, of course).

The complement of (36.9) is $R^T - A = [x \in R^T: (x_{t_1}, \dots, x_{t_k}) \in R^k - H]$, and so \mathcal{R}_0^T is closed under complementation. Suppose that A is given by (36.9) and B is given by

$$(36.10) \quad B = [x \in R^T: (x_{s_1}, \dots, x_{s_j}) \in I],$$

where $I \in \mathcal{R}^j$. Let (u_1, \dots, u_m) be an m -tuple containing all the t_α and all the s_β . Now (t_1, \dots, t_k) must be the initial segment of some permutation of (u_1, \dots, u_m) , and if ψ is as in (36.6) and $H' = \psi^{-1}H$, then $H' \in \mathcal{R}^m$ and A is

given by

$$(36.11) \quad A = [x \in R^T: (x_{u_1}, \dots, x_{u_m}) \in H']$$

as well as by (36.9). Similarly, B can be put in the form

$$(36.12) \quad B = [x \in R^T: (x_{u_1}, \dots, x_{u_m}) \in I'],$$

where $I' \in \mathcal{R}^m$. But then

$$(36.13) \quad A \cup B = [x \in R^T: (x_{u_1}, \dots, x_{u_m}) \in H' \cup I'].$$

Since $H' \cup I' \in \mathcal{R}^m$, $A \cup B$ is a cylinder. This proves that \mathcal{R}_0^T is a field such that $\mathcal{R}^T = \sigma(\mathcal{R}_0^T)$.

The Z_t are measurable functions on the measurable space (R^T, \mathcal{R}^T) . If P is a probability measure on \mathcal{R}^T , then $[Z_t: t \in T]$ is a stochastic process on (R^T, \mathcal{R}^T, P) , the *coordinate-variable process*.

Kolmogorov's Existence Theorem

The existence theorem can be stated two ways:

Theorem 36.1. *If $\mu_{t_1 \dots t_k}$ are a system of distributions satisfying the consistency conditions (36.2) and (36.3), then there is a probability measure P on \mathcal{R}^T such that the coordinate-variable process $[Z_t: t \in T]$ on (R^T, \mathcal{R}^T, P) has the $\mu_{t_1 \dots t_k}$ as its finite-dimensional distributions.*

Theorem 36.2. *If $\mu_{t_1 \dots t_k}$ are a system of distributions satisfying the consistency conditions (36.2) and (36.3), then there exists on some probability space (Ω, \mathcal{F}, P) a stochastic process $[X_t: t \in T]$ having the $\mu_{t_1 \dots t_k}$ as its finite-dimensional distributions.*

For many purposes the underlying probability space is irrelevant, the joint distributions of the variables in the process being all that matters, so that the two theorems are equally useful. As a matter of fact, they are equivalent anyway. Obviously, the first implies the second. To prove the converse, suppose that the process $[X_t: t \in T]$ on (Ω, \mathcal{F}, P) has finite-dimensional distributions $\mu_{t_1 \dots t_k}$, and define a map $\xi: \Omega \rightarrow R^T$ by the requirement

$$(36.14) \quad Z_t(\xi(\omega)) = X_t(\omega), \quad t \in T.$$

For each ω , $\xi(\omega)$ is an element of R^T , a real function on T , and the

requirement is that $X_t(\omega)$ be its value at t . Clearly,

$$\begin{aligned}
 (36.15) \quad & \xi^{-1} \left[x \in R^T : (Z_{t_1}(x), \dots, Z_{t_k}(x)) \in H \right] \\
 &= \left[\omega \in \Omega : (Z_{t_1}(\xi(\omega)), \dots, Z_{t_k}(\xi(\omega))) \in H \right] \\
 &= \left[\omega \in \Omega : (X_{t_1}(\omega), \dots, X_{t_k}(\omega)) \in H \right];
 \end{aligned}$$

since the X_t are random variables, measurable \mathcal{F} , this set lies in \mathcal{F} if $H \in \mathcal{R}^k$. Thus $\xi^{-1}A \in \mathcal{F}$ for $A \in \mathcal{R}_0^T$, and so (Theorem 13.1) ξ is measurable $\mathcal{F}/\mathcal{R}^T$. By (36.15) and the assumption that $[X_t: t \in T]$ has finite-dimensional distributions $\mu_{t_1 \dots t_k}$, $P\xi^{-1}$ (see (13.7)) satisfies

$$\begin{aligned}
 (36.16) \quad & P\xi^{-1} \left[x \in R^T : (Z_{t_1}(x), \dots, Z_{t_k}(x)) \in H \right] \\
 &= P \left[\omega \in \Omega : (X_{t_1}(\omega), \dots, X_{t_k}(\omega)) \in H \right] = \mu_{t_1 \dots t_k}(H).
 \end{aligned}$$

Thus the coordinate-variable process $[Z_t: t \in T]$ on $(R^T, \mathcal{R}^T, P\xi^{-1})$ also has finite-dimensional distributions $\mu_{t_1 \dots t_k}$.

Therefore, to prove either of the two versions of Kolmogorov's existence theorem is to prove the other one as well.

Example 36.1. Suppose that T is finite, say $T = \{1, 2, \dots, k\}$. Then (R^T, \mathcal{R}^T) is (R^k, \mathcal{R}^k) , and taking $P = \mu_{1, 2, \dots, k}$ satisfies the requirements of Theorem 36.1. ■

Example 36.2. Suppose that $T = \{1, 2, \dots\}$ and

$$(36.17) \quad \mu_{t_1 \dots t_k} = \mu_{t_1} \times \cdots \times \mu_{t_k},$$

where μ_1, μ_2, \dots are probability distributions on the line. The consistency conditions are easily checked, and the probability measure P guaranteed by Theorem 36.1 is *product measure* on the product space (R^T, \mathcal{R}^T) . But by Theorem 20.4 there exists on some (Ω, \mathcal{F}, P) an independent sequence X_1, X_2, \dots of random variables with respective distributions μ_1, μ_2, \dots ; then (36.17) is the distribution of $(X_{t_1}, \dots, X_{t_k})$. For the special case (36.17), Theorem 36.2 (and hence Theorem 36.1) was thus proved in Section 20. The existence of independent sequences with prescribed distributions was the measure-theoretic basis of all the probabilistic developments in Chapters 4, 5, and 6: even dependent processes like the Poisson were constructed from independent sequences. The existence of independent sequences can also be made the basis of a proof of Theorems 36.1 and 36.2 in their full generality; see the second proof below. ■

Example 36.3. The preceding example has an analogue in the space S^∞ of sequences (2.15). Here the finite set S plays the role of R^1 , the $z_n(\cdot)$ are analogues of the $Z_n(\cdot)$, and the product measure defined by (2.21) is the analogue of the product measure specified by (36.17) with $\mu_i = \mu$. See also Example 24.2. The theory for S^∞ is simple because S is finite: see Theorem 2.3 and the lemma it depends on. ■

Example 36.4. If T is a subset of the line, it is convenient to use the order structure of the line and take the $\mu_{s_1 \dots s_k}$ to be specified initially only for k -tuples (s_1, \dots, s_k) that are in increasing order:

$$(36.18) \quad s_1 < s_2 < \dots < s_k.$$

It is natural for example to specify the finite-dimensional distributions for the Poisson processes for increasing sequences of time points alone; see (23.27).

Assume that the $\mu_{s_1 \dots s_k}$ for k -tuples satisfying (36.18) have the consistency property

$$(36.19) \quad \mu_{s_1 \dots s_{i-1} s_{i+1} \dots s_k} (H_1 \times \dots \times H_{i-1} \times H_{i+1} \times \dots \times H_k) \\ = \mu_{s_1 \dots s_k} (H_1 \times \dots \times H_{i-1} \times R^1 \times H_{i+1} \times \dots \times H_k).$$

For given s_1, \dots, s_k satisfying (36.18), take $(X_{s_1}, \dots, X_{s_k})$ to have distribution $\mu_{s_1 \dots s_k}$. If t_1, \dots, t_k is a permutation of s_1, \dots, s_k , take $\mu_{t_1 \dots t_k}$ to be the distribution of $(X_{t_1}, \dots, X_{t_k})$:

$$(36.20) \quad \mu_{t_1 \dots t_k} (H_1 \times \dots \times H_k) = P[X_{t_i} \in H_i, i \leq k].$$

This unambiguously defines a collection of finite-dimensional distributions. Are they consistent?

If $t_{\pi 1}, \dots, t_{\pi k}$ is a permutation of t_1, \dots, t_k , then it is also a permutation of s_1, \dots, s_k , and by the definition (36.20), $\mu_{t_{\pi 1} \dots t_{\pi k}}$ is the distribution of $(X_{t_{\pi 1}}, \dots, X_{t_{\pi k}})$, which immediately gives (36.2), the first of the consistency conditions. Because of (36.19), $\mu_{s_1 \dots s_{i-1} s_{i+1} \dots s_k}$ is the distribution of $(X_{s_1}, \dots, X_{s_{i-1}}, X_{s_{i+1}}, \dots, X_{s_k})$, and if $t_k = s_i$, then t_1, \dots, t_{k-1} is a permutation of $s_1, \dots, s_{i-1}, s_{i+1}, \dots, s_k$, which are in increasing order. By the definition (36.20) applied to t_1, \dots, t_{k-1} it therefore follows that $\mu_{t_1 \dots t_{k-1}}$ is the distribution of $(X_{t_1}, \dots, X_{t_{k-1}})$. But this gives (36.3), the second of the consistency conditions.

It will therefore follow from the existence theorem that if $T \subset R^1$ and $\mu_{s_1 \dots s_k}$ is defined for all k -tuples in increasing order, and if (36.19) holds, then there exists a stochastic process $[X_t; t \in T]$ satisfying (36.1) for increasing t_1, \dots, t_k . ■

Two proofs of Kolmogorov's existence theorem will be given. The first is based on the extension theorem of Section 3.

FIRST PROOF OF KOLMOGOROV'S THEOREM. Consider the first formulation, Theorem 36.1. If A is the cylinder (36.9), define

$$(36.21) \quad P(A) = \mu_{t_1 \dots t_k}(H).$$

This gives rise to the question of consistency because A will have other representations as a cylinder. Suppose, in fact, that A coincides with the cylinder B defined by (36.10). As observed before, if (u_1, \dots, u_m) contains all the t_α and s_β , A is also given by (36.11), where $H' = \psi^{-1}H$ and ψ is defined in (36.6). Since the consistency conditions (36.2) and (36.3) imply the more general one (36.7), $P(A) = \mu_{t_1 \dots t_k}(H) = \mu_{u_1 \dots u_m}(H')$. Similarly, (36.10) has the form (36.12), and $P(B) = \mu_{s_1 \dots s_j}(I) = \mu_{u_1 \dots u_m}(I')$. Since the u_γ are distinct, for any real numbers z_1, \dots, z_m there are points x of R^T for which $(x_{u_1}, \dots, x_{u_m}) = (z_1, \dots, z_m)$. From this it follows that if the cylinders (36.11) and (36.12) coincide, then $H' = I'$. Hence $A = B$ implies that $P(A) = \mu_{u_1 \dots u_m}(H') = \mu_{u_1 \dots u_m}(I') = P(B)$, and the definition (36.21) is indeed consistent.

Now consider disjoint cylinders A and B . As usual, the index sets may be taken identical. Assume then that A is given by (36.11) and B by (36.12), so that (36.13) holds. If $H' \cap I'$ were nonempty, then $A \cap B$ would be nonempty as well. Therefore, $H' \cap I' = \emptyset$, and

$$\begin{aligned} P(A \cup B) &= \mu_{u_1 \dots u_m}(H' \cup I') \\ &= \mu_{u_1 \dots u_m}(H') + \mu_{u_1 \dots u_m}(I') = P(A) + P(B). \end{aligned}$$

Therefore, P is finitely additive on \mathcal{R}_0^T . Clearly, $P(R^T) = 1$.

Suppose that P is shown to be countably additive on \mathcal{R}_0^T . By Theorem 3.1, P will then extend to a probability measure on \mathcal{R}^T . By the way P was defined on \mathcal{R}_0^T ,

$$(36.22) \quad P\left[x \in R^T: (Z_{t_1}(x), \dots, Z_{t_k}(x)) \in H\right] = \mu_{t_1 \dots t_k}(H),$$

and therefore the coordinate process $[Z_t: t \in T]$ will have the required finite-dimensional distributions.

It suffices, then, to prove P countably additive on \mathcal{R}_0^T , and this will follow if $A_n \in \mathcal{R}_0^T$ and $A_n \downarrow \emptyset$ together imply $P(A_n) \downarrow 0$ (see Example 2.10). Suppose that $A_1 \supset A_2 \supset \dots$ and that $P(A_n) \geq \epsilon > 0$ for all n . The problem is to show that $\bigcap_n A_n$ must be nonempty. Since $A_n \in \mathcal{R}_0^T$, and since the index set involved in the specification of a cylinder can always be permuted and

expanded, there exists a sequence t_1, t_2, \dots of points in T for which

$$A_n = [x \in R^T: (x_{t_1}, \dots, x_{t_n}) \in H_n],$$

where[†] $H_n \in \mathcal{A}^n$.

Of course, $P(A_n) = \mu_{t_1} \dots \mu_{t_n}(H_n)$. By Theorem 12.3 (regularity), there exists inside H_n a compact set K_n such that $\mu_{t_1} \dots \mu_{t_n}(H_n - K_n) < \epsilon / 2^{n+1}$. If $B_n = [x \in R^T: (x_{t_1}, \dots, x_{t_n}) \in K_n]$, then $P(A_n - B_n) < \epsilon / 2^{n+1}$. Put $C_n = \bigcap_{k=1}^n B_k$. Then $C_n \subset B_n \subset A_n$ and $P(A_n - C_n) < \epsilon / 2$, so that $P(C_n) > \epsilon / 2 > 0$. Therefore, $C_n \subset C_{n-1}$ and C_n is nonempty.

Choose a point $x^{(n)}$ of R^T in C_n . If $n \geq k$, then $x^{(n)} \in C_n \subset C_k \subset B_k$ and hence $(x_{t_1}^{(n)}, \dots, x_{t_k}^{(n)}) \in K_k$. Since K_k is bounded, the sequence $\{x_{t_k}^{(1)}, x_{t_k}^{(2)}, \dots\}$ is bounded for each k . By the diagonal method [A14] select an increasing sequence n_1, n_2, \dots of integers such that $\lim_i x_{t_k}^{(n_i)}$ exists for each k . There is in R^T some point x whose t_k th coordinate is this limit for each k . But then, for each k , $(x_{t_1}, \dots, x_{t_k})$ is the limit as $i \rightarrow \infty$ of $(x_{t_1}^{(n_i)}, \dots, x_{t_k}^{(n_i)})$ and hence lies in K_k . But that means that x itself lies in B_k and hence in A_k . Thus $x \in \bigcap_{k=1}^\infty A_k$, which completes the proof.[‡] ■

The second proof of Kolmogorov's theorem goes in two stages, first for countable T , then for general T .^{*}

SECOND PROOF FOR COUNTABLE T . The result for countable T will be proved in its second formulation, Theorem 36.2. It is no restriction to enumerate T as $\{t_1, t_2, \dots\}$ and then to identify t_n with n ; in other words, it is no restriction to assume that $T = \{1, 2, \dots\}$. Write μ_n in place of $\mu_{1,2,\dots,n}$.

By Theorem 20.4 there exists on a probability space (Ω, \mathcal{F}, P) (which can be taken to be the unit interval) an independent sequence U_1, U_2, \dots of random variables each uniformly distributed over $(0, 1)$. Let F_1 be the distribution function corresponding to μ_1 . If the "inverse" g_1 of F_1 is defined over $(0, 1)$ by $g_1(s) = \inf\{x: s \leq F_1(x)\}$, then $X_1 = g_1(U_1)$ has distribution μ_1 by the usual argument: $P[g_1(U_1) \leq x] = P[U_1 \leq F_1(x)] = F_1(x)$.

The problem is to construct X_2, X_3, \dots inductively in such a way that

$$(36.23) \quad X_k = h_k(U_1, \dots, U_k)$$

for a Borel function h_k and (X_1, \dots, X_n) has the distribution μ_n . Assume that X_1, \dots, X_{n-1} have been defined ($n \geq 2$): they have joint distribution μ_{n-1} and (36.23) holds for $k \leq n-1$. The idea now is to construct an appropriate conditional distribution function $F_n(x|x_1, \dots, x_{n-1})$; here $F_n(x|X_1(\omega), \dots, X_{n-1}(\omega))$ will have the value $P[X_n \leq x | X_1, \dots, X_{n-1}]_\omega$ would have if X_n were already defined. If $g_n(\cdot | x_1, \dots, x_{n-1})$

[†]In general, A_n will involve indices t_1, \dots, t_{a_n} , where $a_1 < a_2 < \dots$. For notational simplicity a_n is taken as n . As a matter of fact, this can be arranged anyway: Take $A'_{a_n} = A_n$, $A'_k = [x: (x_{t_1}, \dots, x_{t_k}) \in R^k] = R^T$ for $k < a_1$, and $A'_k = [x: (x_{t_1}, \dots, x_{t_k}) \in H_n \times R^{k-a_n}] = A_n$ for $a_n < k < a_{n+1}$. Now relabel A'_n as A_n .

[‡]The last part of the argument is, in effect, the proof that a countable product of compact sets is compact.

^{*}This second proof, which may be omitted, uses the conditional-probability theory of Section 33.

is the "inverse" function, then $X_n(\omega) = g_n(U_n(\omega) | X_1(\omega), \dots, X_{n-1}(\omega))$ will by the usual argument have the right conditional distribution given X_1, \dots, X_{n-1} , so that $(X_1, \dots, X_{n-1}, X_n)$ will have the right distribution over R^n .

To construct the conditional distribution function, apply Theorem 33.3 in $(R^n, \mathcal{R}^n, \mu_n)$ to get a conditional distribution of the last coordinate of (x_1, \dots, x_n) given the first $n-1$ of them. This will have (Theorem 20.1) the form $\nu(H; x_1, \dots, x_{n-1})$; it is a probability measure as H varies over \mathcal{R}^1 , and

$$\begin{aligned} & \int_{(x_1, \dots, x_{n-1}) \in M} \nu(H; x_1, \dots, x_{n-1}) d\mu_n(x_1, \dots, x_n) \\ &= \mu_n[x \in R^n: (x_1, \dots, x_{n-1}) \in M, x_n \in H]. \end{aligned}$$

Since the integrand involves only x_1, \dots, x_{n-1} , and since μ_n by consistency projects to μ_{n-1} under the map $(x_1, \dots, x_n) \rightarrow (x_1, \dots, x_{n-1})$, a change of variable gives

$$\begin{aligned} & \int_M \nu(H; x_1, \dots, x_{n-1}) d\mu_{n-1}(x_1, \dots, x_{n-1}) \\ &= \mu_n[x \in R^n: (x_1, \dots, x_{n-1}) \in M, x_n \in H]. \end{aligned}$$

Define $F_n(x | x_1, \dots, x_{n-1}) = \nu((-\infty, x]; x_1, \dots, x_{n-1})$. Then $F_n(\cdot | x_1, \dots, x_{n-1})$ is a probability distribution function over the line, $F_n(x | \cdot)$ is a Borel function over R^{n-1} , and

$$\begin{aligned} & \int_M F_n(x | x_1, \dots, x_{n-1}) d\mu_{n-1}(x_1, \dots, x_{n-1}) \\ &= \mu_n[x \in R^n: (x_1, \dots, x_{n-1}) \in M, x_n \leq x]. \end{aligned}$$

Put $g_n(u | x_1, \dots, x_{n-1}) = \inf\{x: u \leq F_n(x | x_1, \dots, x_{n-1})\}$ for $0 < u < 1$. Since $F_n(x | x_1, \dots, x_{n-1})$ is nondecreasing and right-continuous in x , $g_n(u | x_1, \dots, x_{n-1}) \leq x$ if and only if $u \leq F_n(x | x_1, \dots, x_{n-1})$. Set $X_n = g_n(U_n | X_1, \dots, X_{n-1})$. Since (X_1, \dots, X_{n-1}) has distribution μ_{n-1} and by (36.23) is independent of U_n , an application of (20.30) gives

$$\begin{aligned} & P[(X_1, \dots, X_{n-1}) \in M, X_n \leq x] \\ &= P[(X_1, \dots, X_{n-1}) \in M, U_n \leq F_n(x | X_1, \dots, X_{n-1})] \\ &= \int_M P[U_n \leq F_n(x | x_1, \dots, x_{n-1})] d\mu_{n-1}(x_1, \dots, x_{n-1}) \\ &= \int_M F_n(x | x_1, \dots, x_{n-1}) d\mu_{n-1}(x_1, \dots, x_{n-1}) \\ &= \mu_n[x \in R^n: (x_1, \dots, x_{n-1}) \in M, x_n \leq x]. \end{aligned}$$

Thus (X_1, \dots, X_n) has distribution μ_n . Note that X_n , as a function of X_1, \dots, X_{n-1} and U_n , is a function of U_1, \dots, U_n because (36.23) was assumed to hold for $k < n$. Hence (36.23) holds for $k = n$ as well. ■

SECOND PROOF FOR GENERAL T . Consider (R^T, \mathcal{R}^T) once again. If $S \subset T$, let $\mathcal{F}_S = \sigma[Z_t: t \in S]$. Then $\mathcal{F}_S \subset \mathcal{F}_T = \mathcal{R}^T$.

Suppose that S is countable. By the case just treated, there exists a process $[X_t: t \in S]$ on some (Ω, \mathcal{F}, P) —the space and the process depend on S —such that $(X_{t_1}, \dots, X_{t_k})$ has distribution μ_{t_1, \dots, t_k} for every k -tuple (t_1, \dots, t_k) from S . Define a map $\xi: \Omega \rightarrow R^T$ by requiring that

$$Z_t(\xi(\omega)) = \begin{cases} X_t(\omega) & \text{if } t \in S, \\ 0 & \text{if } t \notin S. \end{cases}$$

Now (36.15) holds as before if t_1, \dots, t_k all lie in S , and so ξ is measurable $\mathcal{F}/\mathcal{F}_S$. Further, (36.16) holds for t_1, \dots, t_k in S . Put $P_S = P\xi^{-1}$ on \mathcal{F}_S . Then P_S is a probability measure on (R^T, \mathcal{F}_S) , and

$$(36.24) \quad P_S[x \in R^T: (Z_{t_1}(x), \dots, Z_{t_k}(x)) \in H] = \mu_{t_1, \dots, t_k}(H)$$

if $H \in \mathcal{R}^k$ and t_1, \dots, t_k all lie in S . (The various spaces (Ω, \mathcal{F}, P) and processes $[X_t: t \in S]$ now become irrelevant.)

If $S_0 \subset S$, and if A is a cylinder (36.9) for which the t_1, \dots, t_k lie in S_0 , then $P_{S_0}(A)$ and $P_S(A)$ coincide, their common value being $\mu_{t_1, \dots, t_k}(H)$. Since these cylinders generate \mathcal{F}_{S_0} , $P_{S_0}(A) = P_S(A)$ for all A in \mathcal{F}_{S_0} . If A lies both in \mathcal{F}_{S_1} and \mathcal{F}_{S_2} , then $P_{S_1}(A) = P_{S_1 \cup S_2}(A) = P_{S_2}(A)$. Thus $P(A) = P_S(A)$ consistently defines a set function on the class $\bigcup_S \mathcal{F}_S$, the union extending over the countable subsets S of T . If A_n lies in this union and $A_n \in \mathcal{F}_{S_n}$ (S_n countable), then $S = \bigcup_n S_n$ is countable and $\bigcup_n A_n$ lies in \mathcal{F}_S . Thus $\bigcup_S \mathcal{F}_S$ is a σ -field and so must coincide with \mathcal{R}^T . Therefore, P is a probability measure on \mathcal{R}^T , and by (36.24) the coordinate process has under P the required finite-dimensional distributions. ■

The Inadequacy of \mathcal{R}^T

Theorem 36.3. *Let $[X_t: t \in T]$ be a family of real functions on Ω .*

(i) *If $A \in \sigma[X_t: t \in T]$ and $\omega \in A$, and if $X_t(\omega) = X_t(\omega')$ for all $t \in T$, then $\omega' \in A$.*

(ii) *If $A \in \sigma[X_t: t \in T]$, then $A \in \sigma[X_t: t \in S]$ for some countable subset S of T .*

PROOF. Define $\xi: \Omega \rightarrow R^T$ by $Z_t(\xi(\omega)) = X_t(\omega)$. Let $\mathcal{F} = \sigma[X_t: t \in T]$. By (36.15), ξ is measurable $\mathcal{F}/\mathcal{R}^T$ and hence \mathcal{F} contains the class $[\xi^{-1}M: M \in \mathcal{R}^T]$. The latter class is a σ -field, however, and by (36.15) it contains the sets $[\omega \in \Omega: (X_{t_1}(\omega), \dots, X_{t_k}(\omega)) \in H]$, $H \in \mathcal{R}^k$, and hence contains the σ -field \mathcal{F} they generate. Therefore

$$(36.25) \quad \sigma[X_t: t \in T] = [\xi^{-1}M: M \in \mathcal{R}^T].$$

This is an infinite-dimensional analogue of Theorem 20.1(i).

As for (i), the hypotheses imply that $\omega \in A = \xi^{-1}M$ and $\xi(\omega) = \xi(\omega')$, so that $\omega' \in A$ certainly follows.

For $S \subset T$, let $\mathcal{F}_S = \sigma[X_t: t \in S]$; (ii) says that $\mathcal{F} = \mathcal{F}_T$ coincides with $\mathcal{G} = \bigcup_S \mathcal{F}_S$, the union extending over the countable subsets S of T . If A_1, A_2, \dots lie in \mathcal{G} , A_n lies in \mathcal{F}_{S_n} for some countable S_n , and so $\bigcup_n A_n$ lies in \mathcal{G} because it lies in \mathcal{F}_S for $S = \bigcup_n S_n$. Thus \mathcal{G} is a σ -field, and since it contains the sets $[X_t \in H]$, it contains the σ -field \mathcal{F} they generate. (This part of the argument was used in the second proof of the existence theorem.) ■

From this theorem it follows that various important sets lie outside the class \mathcal{R}^T . Suppose that $T = [0, \infty)$. Of obvious interest is the subset C of R^T consisting of the functions continuous over $[0, \infty)$. But C is not in \mathcal{R}^T . For suppose it were. By part (ii) of the theorem (let $\Omega = R^T$ and put $[Z_t: t \in T]$ in the role of $[X_t: t \in T]$), C would lie in $\sigma[Z_t: t \in S]$ for some countable $S \subset [0, \infty)$. But then by part (i) of the theorem (let $\Omega = R^T$ and put $[Z_t: t \in S]$ in the role of $[X_t: t \in T]$), if $x \in C$ and $Z_t(x) = Z_t(y)$ for all $t \in S$, then $y \in C$. From the assumption that C lies in \mathcal{R}^T thus follows the existence of a countable set S such that, if $x \in C$ and $x(t) = y(t)$ for all t in S , then $y \in C$. But whatever countable set S may be, for every continuous x there obviously exist functions y that have discontinuities but agree with x on S . Therefore, C cannot lie in \mathcal{R}^T .

What the argument shows is this: A set A in R^T cannot lie in \mathcal{R}^T unless there exists a countable subset S of T with the property that, if $x \in A$ and $x(t) = y(t)$ for all t in S , then $y \in A$. Thus A cannot lie in \mathcal{R}^T if it effectively involves all the points t in the sense that, for each x in A and each t in T , it is possible to move x out of A by changing its value at t alone. And C is such a set. For another, consider the set of functions x over $T = [0, \infty)$ that are nondecreasing and assume as values $x(t)$ only nonnegative integers:

$$(36.26) \quad [x \in R^{[0, \infty)}: x(s) \leq x(t), x \leq t; x(t) \in \{0, 1, \dots\}, t \geq 0].$$

This, too, lies outside \mathcal{R}^T .

In Section 23 the Poisson process was defined as follows: Let X_1, X_2, \dots be independent and identically distributed with the exponential distribution (the probability space Ω on which they are defined may by Theorem 20.4 be taken to be the unit interval with Lebesgue measure). Put $S_0 = 0$ and $S_n = X_1 + \dots + X_n$. If $S_n(\omega) < S_{n+1}(\omega)$ for $n \geq 0$ and $S_n(\omega) \rightarrow \infty$, put $N(t, \omega) = N_t(\omega) = \max[n: S_n(\omega) \leq t]$ for $t \geq 0$; otherwise, put $N(t, \omega) = N_t(\omega) = 0$ for $t \geq 0$. Then the stochastic process $[N_t: t \geq 0]$ has the finite-dimensional distributions described by the equations (23.27). The function $N(\cdot, \omega)$ is the *path function* or *sample function*[†] corresponding to ω , and by the construction every path function lies in the set (36.26). This is a good thing if the

[†]Other terms are *realization* of the process and *trajectory*.

process is to be a model for, say, calls arriving at a telephone exchange: The sample path represents the history of the calls, its value at t being the number of arrivals up to time t , and so it ought to be nondecreasing and integer-valued.

According to Theorem 36.1, there exists a measure P on R^T for $T = [0, \infty)$ such that the coordinate process $[Z_t: t \geq 0]$ on (R^T, \mathcal{R}^T, P) has the finite-dimensional distributions of the Poisson process. This time does the path function $Z(\cdot, x)$ lie in the set (36.26) with probability 1? Since $Z(\cdot, x)$ is just x itself, the question is whether the set (36.26) has P -measure 1. But this set does not lie in \mathcal{R}^T , and so it has no measure at all.

An application of Kolmogorov's existence theorem will always yield a stochastic process with prescribed finite-dimensional distributions, but the process may lack certain path-function properties that it is reasonable to require of it as a model for some natural phenomenon. The special construction of Section 23 gets around this difficulty for the Poisson process, and in the next section a special construction will yield a model for Brownian motion with continuous paths. Section 38 treats a general method for producing stochastic processes that have prescribed finite-dimensional distributions and at the same time have path functions with desirable regularity properties.

A Return to Ergodic Theory*

Write $R^\infty, \mathcal{R}_0^\infty, \mathcal{R}^\infty$ for $R^T, \mathcal{R}_0^T, \mathcal{R}^T$ in the case where the index set $\{0, \pm 1, \pm 2, \dots\}$ consists of all the integers. Then R^∞ is analogous to S^∞ (Sections 2 and 24), except that here the sequences are doubly infinite:

$$x = (\dots, Z_{-1}(x), Z_0(x), Z_1(x), \dots).$$

Let T (not an index set) denote the *shift*: $Z_k(Tx) = Z_{k+1}(x)$, $k = 0, \pm 1, \dots$. This is like the shift in Section 24. Since $A \in \mathcal{R}_0^\infty$ implies $T^{-1}A \in \mathcal{R}_0^\infty$, T is measurable $\mathcal{R}^\infty / \mathcal{R}^\infty$. Clearly, it is invertible.

For a stochastic process $X = (\dots, X_{-1}, X_0, X_1, \dots)$ on (Ω, \mathcal{F}, P) , define $\xi: \Omega \rightarrow R^\infty$ by (36.14): $\xi\omega = X(\omega) = (\dots, X_{-1}(\omega), X_0(\omega), X_1(\omega), \dots)$. The measure $P\xi^{-1} = PX^{-1}$ on $(R^\infty, \mathcal{R}^\infty)$ can be viewed as the *distribution* of X . Suppose that X is *stationary* in the sense that, for each $k \geq 1$ and $H \in \mathcal{R}^k$, $P[(X_{n+1}, \dots, X_{n+k}) \in H]$ is the same for all $n = 0, \pm 1, \dots$. Then the shift preserves $P\xi^{-1}$ (use (36.16) and Lemma 1, p. 311). The process X is defined to be *ergodic* if under $P\xi^{-1}$ the shift is ergodic in the sense of Section 24.

In the ergodic case, it follows by the ergodic theorem that

$$(36.27) \quad \frac{1}{n} \sum_{k=1}^n f(T^k x) \rightarrow \int_{R^\infty} f(x) P\xi^{-1}(dx)$$

*This topic, which requires Section 24, may be omitted.

on a set of $P\xi^{-1}$ -measure 1, provided f is measurable \mathcal{R}^∞ and integrable. Carry (36.27) back to (Ω, \mathcal{F}, P) by the inverse set mapping ξ^{-1} . Then

$$(36.28) \quad \frac{1}{n} \sum_{k=1}^n f(\dots, X_{k-1}, X_k, X_{k+1}, \dots) \rightarrow E[f(\dots, X_{-1}, X_0, X_1, \dots)]$$

with probability 1: (36.28) holds at ω if and only if (36.27) holds at $x = \xi\omega = X(\omega)$. It is understood that on the left in (36.28), X_k is the center coordinate (the 0th coordinate) of the argument of f , and on the right, X_0 is the center coordinate: *For stationary, ergodic X and integrable f , (36.28) holds with probability 1.*

If the X_k are independent, then the Z_k are independent under $P\xi^{-1}$. In this case, $\lim_n P\xi^{-1}(A \cap T^{-n}B) = P\xi^{-1}(A)P\xi^{-1}(B)$ for A and B in \mathcal{R}_0^∞ , because for large enough n the cylinders A and $T^{-n}B$ depend on disjoint sets of time indices and hence are independent. But then it follows by approximation (Corollary 1 to Theorem 11.4) that the same limit holds for all A and B in \mathcal{R}^∞ . But for invariant B , this implies $P\xi^{-1}(B^c \cap B) = P\xi^{-1}(B^c)P\xi^{-1}(B)$, so that $P\xi^{-1}(B)$ is 0 or 1, and the shift is ergodic under $P\xi^{-1}$: *If X is stationary and independent, then it is ergodic.*

If f depends on just one coordinate of x , then (36.28) is in the independent case a consequence of the strong law of large numbers, Theorem 22.1. But (36.28) follows by the ergodic theorem even if f involves all the coordinates in some complicated way.

Consider now a measurable real function ϕ on R^∞ . Define $\psi: R^\infty \rightarrow R^\infty$ by

$$\psi(x) = (\dots, \phi(T^{-1}x), \phi(x), \phi(Tx), \dots);$$

here $\phi(x)$ is the center coordinate: $Z_k(\psi(x)) = \phi(T^k x)$. It is easy to show that ψ is measurable $\mathcal{R}^\infty/\mathcal{R}^\infty$ and commutes with the shift in the sense of Example 24.6. Therefore, T preserves $P\xi^{-1}\psi^{-1}$ if it preserves $P\xi^{-1}$, and it is ergodic under $P\xi^{-1}\psi^{-1}$ if it is ergodic under $P\xi^{-1}$.

This translates immediately into a result on stochastic processes. Define $Y = (\dots, Y_{-1}, Y_0, Y_1, \dots)$ in terms of X by

$$(36.29) \quad Y_n = \phi(\dots, X_{n-1}, X_n, X_{n+1}, \dots),$$

that is to say, $Y(\omega) = \psi(X(\omega)) = \psi\xi\omega$. Since $P\xi^{-1}$ is the distribution of X , $P\xi^{-1}\psi^{-1} = P(\psi\xi)^{-1} = PY^{-1}$ is the distribution of Y :

Theorem 36.4. *If X is stationary and ergodic, in particular if the X_n are independent and identically distributed, then Y as defined by (36.29) is stationary and ergodic.*

This theorem fails if Y is not defined in terms of X in a time-invariant way—if the ϕ in (36.29) is not the same for all n : If $\phi_n(x) = Z_{-n}(x)$ and ϕ is replaced by ϕ_n in (36.29), then $Y_n \equiv X_0$; in this case Y happens to be stationary, but it is not ergodic if the distribution of X_0 does not concentrate at a single point.

Example 36.5. *The autoregressive model.* Let $\phi(x) = \sum_{k=0}^\infty \beta^k Z_{-k}(x)$ on the set where the series converges, and take $\phi(x) = 0$ elsewhere. Suppose that $|\beta| < 1$ and that the X_n are independent and identically distributed with finite second moments. Then by Theorem 22.6, $Y_n = \sum_{k=0}^\infty \beta^k X_{n-k}$ converges with probability 1, and by Theorem 36.4, the process Y is ergodic. Note that $Y_{n+1} = \beta Y_n + X_{n+1}$ and that X_{n+1} is independent of Y_n . This is the linear autoregressive model of order 1. ■

The Hewitt–Savage Theorem*

Change notation: Let $(R^\infty, \mathcal{R}^\infty)$ be the product space with $\{1, 2, \dots\}$ as the index set, the space of one-sided sequences. Let P be a probability measure on \mathcal{R}^∞ . If the coordinate variables Z_n are independent under P , then by Theorem 22.3, $P(A)$ is 0 or 1 for each A in the tail σ -field \mathcal{T} . If the Z_n are also identically distributed under P , a stronger result holds.

Let \mathcal{S}_n be the class of \mathcal{R}^∞ -sets A that are invariant under permutations of the first n coordinates: if π is a permutation of $\{1, \dots, n\}$, then x lies in A if and only if $(Z_{\pi_1}(x), \dots, Z_{\pi_n}(x), Z_{n+1}(x), \dots)$ does. Then \mathcal{S}_n is a σ -field. Let $\mathcal{S} = \bigcap_{n=1}^\infty \mathcal{S}_n$ be the σ -field of \mathcal{R}^∞ -sets invariant under all finite permutations of coordinates. Then \mathcal{S} is larger than \mathcal{T} , since, for example, the x -set where $\sum_{k=1}^n Z_k(x) > c_n$ infinitely often lies in \mathcal{S} but not in \mathcal{T} .

The *Hewitt–Savage theorem* is a zero–one law for \mathcal{S} in the independent, identically distributed case.

Theorem 36.5. *If the Z_n are independent and identically distributed under P , then $P(A)$ is 0 or 1 for each A in \mathcal{S} .*

PROOF. By Corollary 1 to Theorem 11.4, there are for given A and ϵ an n and a set $U = [(Z_1, \dots, Z_n) \in H]$ ($H \in \mathcal{R}^n$) such that $P(A \Delta U) < \epsilon$. Let $V = [(Z_{n+1}, \dots, Z_{2n}) \in H]$. If the Z_k are independent and identically distributed, then $P(A \Delta U)$ is the same as

$$P([(Z_{n+1}, \dots, Z_{2n}, Z_1, \dots, Z_n, Z_{2n+1}, Z_{2n+2}, \dots) \in A] \\ \Delta [(Z_{n+1}, \dots, Z_{2n}, Z_1, \dots, Z_n) \in H \times R^n]).$$

But if $A \in \mathcal{S}_{2n}$, this is in turn the same as $P(A \Delta V)$. Therefore, $P(A \Delta U) = P(A \Delta V)$.

But then, $P(A \Delta V) < \epsilon$ and $P(A \Delta (U \cap V)) \leq P(A \Delta U) + P(A \Delta V) < 2\epsilon$. Since U and V have the same probability and are independent, it follows that $P(A)$ is within ϵ of $P(U)$ and hence $P^2(A)$ is within 2ϵ of $P^2(U) = P(U)P(V) = P(U \cap V)$, which is in turn within 2ϵ of $P(A)$. Therefore, $|P^2(A) - P(A)| < 4\epsilon$ for all ϵ , and so $P(A)$ must be 0 or 1. ■

PROBLEMS

36.1. ↑ Suppose that $[X_t: t \in T]$ is a stochastic process on (Ω, \mathcal{F}, P) and $A \in \mathcal{F}$. Show that there is a countable subset S of T for which $P[A \| X_t, t \in T] = P[A \| X_t, t \in S]$ with probability 1. Replace A by a random variable and prove a similar result.

36.2. Let T be arbitrary and let $K(s, t)$ be a real function over $T \times T$. Suppose that K is symmetric in the sense that $K(s, t) = K(t, s)$ and nonnegative-definite in the sense that $\sum_{i,j=1}^k K(t_i, t_j) x_i x_j \geq 0$ for $k \geq 1$, t_1, \dots, t_k in T , and x_1, \dots, x_k real. Show that there exists a process $[X_t: t \in T]$ for which $(X_{t_1}, \dots, X_{t_k})$ has the centered normal distribution with covariances $K(t_i, t_j)$, $i, j = 1, \dots, k$.

*This topic may be omitted.

36.3. Let L be a Borel set on the line, let \mathcal{L} consist of the Borel subsets of L , and let L^T consist of all maps from T into L . Define the appropriate notion of cylinder, and let \mathcal{L}^T be the σ -field generated by the cylinders. State a version of Theorem 36.1 for (L^T, \mathcal{L}^T) . Assume T countable, and prove this theorem not by imitating the previous proof but by observing that L^T is a subset of R^T and lies in \mathcal{R}^T .

36.4. Suppose that the random variables X_1, X_2, \dots assume the values 0 and 1 and $P[X_n = 1 \text{ i.o.}] = 1$. Let μ be the distribution over $(0, 1]$ of $\sum_{n=1}^{\infty} X_n / 2^n$. Show that on the unit interval with the measure μ , the digits of the nonterminating dyadic expansion form a stochastic process with the same finite-dimensional distributions as X_1, X_2, \dots .

36.5. 36.3 \uparrow There is an infinite-dimensional version of Fubini's theorem. In the construction in Problem 36.3, let $L = I = (0, 1)$, $T = \{1, 2, \dots\}$, let \mathcal{J} consist of the Borel subsets of I , and suppose that each k -dimensional distribution is the k -fold product of Lebesgue measure over the unit interval. Then I^T is a countable product of copies of $(0, 1)$, its elements are sequences $x = (x_1, x_2, \dots)$ of points of $(0, 1)$, and Kolmogorov's theorem ensures the existence on (I^T, \mathcal{J}^T) of a *product* probability measure π : $\pi[x: x_i \leq \alpha_i, i \leq n] = \alpha_1 \cdots \alpha_n$ for $0 \leq \alpha_i \leq 1$. Let I^n denote the n -dimensional unit cube.

(a) Define $\psi: I^n \times I^T \rightarrow I^T$ by

$$\psi((x_1, \dots, x_n), (y_1, y_2, \dots)) = (x_1, \dots, x_n, y_1, y_2, \dots).$$

Show that ψ is measurable $\mathcal{J}^n \times \mathcal{J}^T / \mathcal{J}^T$ and ψ^{-1} is measurable $\mathcal{J}^T / \mathcal{J}^n \times \mathcal{J}^T$. Show that $\psi^{-1}(\lambda_n \times \pi) = \pi$, where λ_n is n -dimensional Lebesgue measure restricted to I^n .

(b) Let f be a function measurable \mathcal{J}^T and, for simplicity, bounded. Define

$$f_n(x_{n+1}, x_{n+2}, \dots) = \int_0^1 \cdots \int_0^1 f(y_1, \dots, y_n, x_{n+1}, \dots) dy_1 \cdots dy_n;$$

in other words, integrate out the coordinates one by one. Show by Problem 34.18, martingale theory, and the zero-one law that

$$(36.30) \quad f_n(x_{n+1}, x_{n+2}, \dots) \rightarrow \int_{I^T} f(y) \pi(dy)$$

except for x in a set of π -measure 0.

(c) Adopting the point of view of part (a), let $g_n(x_1, \dots, x_n)$ be the result of integrating the variable $(y_{n+1}, y_{n+2}, \dots)$ out (with respect to π) from $f(x_1, \dots, x_n, y_{n+1}, \dots)$. This may suggestively be written as

$$g_n(x_1, \dots, x_n) = \int_0^1 \int_0^1 \cdots f(x_1, \dots, x_n, y_{n+1}, y_{n+2}, \dots) dy_{n+1} dy_{n+2} \cdots$$

Show that $g_n(x_1, \dots, x_n) \rightarrow f(x_1, x_2, \dots)$ except for x in a set of π -measure 0.

36.6. (a) Let T be an interval of the line. Show that \mathcal{R}^T fails to contain the sets of: linear functions, polynomials, constants, nondecreasing functions, functions of bounded variation, differentiable functions, analytic functions, functions con-

tinuous at a fixed t_0 , Borel measurable functions. Show that it fails to contain the set of functions that: vanish somewhere in T , satisfy $x(s) < x(t)$ for some pair with $s < t$, have a local maximum anywhere, fail to have a local maximum.

(b) Let C be the set of continuous functions on $T = [0, \infty)$. Show that $A \in \mathcal{R}^T$ and $A \subset C$ imply that $A = \emptyset$. Show, on the other hand, that $A \in \mathcal{R}^T$ and $C \subset A$ do not imply that $A = R^T$.

36.7. Not all systems of finite-dimensional distributions can be realized by stochastic processes for which Ω is the unit interval. Show that there is on the unit interval with Lebesgue measure no process $[X_t: t \geq 0]$ for which the X_t are independent and assume the values 0 and 1 with probability $\frac{1}{2}$ each. Compare Problem 1.1.

36.8. Here is an application of the existence theorem in which T is not a subset of the line. Let (N, \mathcal{N}, ν) be a measure space, and take T to consist of the \mathcal{N} -sets of finite ν -measure. The problem is to construct a generalized Poisson process, a stochastic process $[X_A: A \in T]$ such that (i) X_A has the Poisson distribution with mean $\nu(A)$ and (ii) X_{A_1}, \dots, X_{A_n} are independent if A_1, \dots, A_n are disjoint. *Hint:* To define the finite-dimensional distributions, generalize this construction: For A, B in T , consider independent random variables Y_1, Y_2, Y_3 having Poisson distributions with means $\nu(A \cap B^c)$, $\nu(A \cap B)$, $\nu(A^c \cap B)$; take $\mu_{A,B}$ to be the distribution of $(Y_1 + Y_2, Y_2 + Y_3)$.

SECTION 37. BROWNIAN MOTION

Definition

A *Brownian motion* or *Wiener process* is a stochastic process $[W_t: t \geq 0]$, on some (Ω, \mathcal{F}, P) , with these three properties:

(i) *The process starts at 0:*

$$(37.1) \quad P[W_0 = 0] = 1.$$

(ii) *The increments are independent: If*

$$(37.2) \quad 0 \leq t_0 \leq t_1 \leq \dots \leq t_k,$$

then

$$(37.3) \quad P[W_{t_i} - W_{t_{i-1}} \in H_i, i \leq k] = \prod_{i \leq k} P[W_{t_i} - W_{t_{i-1}} \in H_i].$$

(iii) *For $0 \leq s < t$ the increment $W_t - W_s$ is normally distributed with mean 0 and variance $t - s$:*

$$(37.4) \quad P[W_t - W_s \in H] = \frac{1}{\sqrt{2\pi(t-s)}} \int_H e^{-x^2/2(t-s)} dx.$$

The existence of such processes will be proved.

Imagine suspended in a fluid a particle bombarded by molecules in thermal motion. The particle will perform a seemingly random movement

first described by the nineteenth-century botanist Robert Brown. Consider a single component of this motion—imagine it projected on a vertical axis—and denote by W_t the height at time t of the particle above a fixed horizontal plane. Condition (i) is merely a convention: the particle starts at 0. Condition (ii) reflects a kind of lack of memory. The displacements $W_{t_1} - W_{t_0}, \dots, W_{t_{k-1}} - W_{t_{k-2}}$ the particle undergoes during the intervals $[t_0, t_1], \dots, [t_{k-2}, t_{k-1}]$ in no way influence the displacement $W_{t_k} - W_{t_{k-1}}$ it undergoes during $[t_{k-1}, t_k]$. Although the future behavior of the particle depends on its present position, it does not depend on how the particle got there. As for (iii), that $W_t - W_s$ has mean 0 reflects the fact that the particle is as likely to go up as to go down—there is no drift. The variance grows as the length of the interval $[s, t]$; the particle tends to wander away from its position at time s , and having done so suffers no force tending to restore it to that position. To Norbert Wiener are due the mathematical foundations of the theory of this kind of random motion.



A Brownian motion path.

The increments of the Brownian motion process are *stationary* in the sense that the distribution of $W_t - W_s$ depends only on the difference $t - s$. Since $W_0 = 0$, the distribution of these increments is described by saying that

W_t is normally distributed with mean 0 and variance t . This implies (37.1). If $0 \leq s \leq t$, then by the independence of the increments, $E[W_s W_t] = E[(W_s(W_t - W_s) + W_s^2)] = E[W_s]E[W_t - W_s] + E[W_s^2] = s$. This specifies all the means, variances, and covariances:

$$(37.5) \quad E[W_t] = 0, \quad E[W_t^2] = t, \quad E[W_s W_t] = \min\{s, t\}.$$

If $0 < t_1 < \cdots < t_k$, the joint density of $(W_{t_1}, W_{t_2} - W_{t_1}, \dots, W_{t_k} - W_{t_{k-1}})$ is by (20.25) the product of the corresponding normal densities. By the Jacobian formula (20.20), $(W_{t_1}, \dots, W_{t_k})$ has density

$$(37.6) \quad f_{t_1, \dots, t_k}(x_1, \dots, x_k) = \prod_{i=1}^k \frac{1}{\sqrt{2\pi(t_i - t_{i-1})}} \exp\left[-\frac{(x_i - x_{i-1})^2}{2(t_i - t_{i-1})}\right],$$

where $t_0 = x_0 = 0$.

Sometimes W_t will be denoted $W(t)$, and its value at ω will be $W(t, \omega)$. The nature of the path functions $W(\cdot, \omega)$ will be of great importance.

The existence of the Brownian motion process follows from Kolmogorov's theorem. For $0 < t_1 < \cdots < t_k$ let μ_{t_1, \dots, t_k} be the distribution in R^k with density (37.6). To put it another way, let μ_{t_1, \dots, t_k} be the distribution of (S_1, \dots, S_k) , where $S_i = X_1 + \cdots + X_i$ and where X_1, \dots, X_k are independent, normally distributed random variables with mean 0 and variances $t_1, t_2 - t_1, \dots, t_k - t_{k-1}$. If $g(x_1, \dots, x_k) = (x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_k)$, then $g(S_1, \dots, S_k) = (S_1, \dots, S_{i-1}, S_{i+1}, \dots, S_k)$ has the distribution prescribed for $\mu_{t_1, \dots, t_{i-1}, t_{i+1}, \dots, t_k}$. This is because $X_i + X_{i+1}$ is normally distributed with mean 0 and variance $t_{i+1} - t_{i-1}$; see Example 20.6. Therefore,

$$(37.7) \quad \mu_{t_1, \dots, t_{i-1}, t_{i+1}, \dots, t_k} = \mu_{t_1, \dots, t_k} g^{-1}.$$

The μ_{t_1, \dots, t_k} defined in this way for increasing, positive t_1, \dots, t_k thus satisfy the conditions for Kolmogorov's existence theorem as modified in Example 36.4; (37.7) is the same thing as (36.19). Therefore, there does exist a process $[W_t: t > 0]$ corresponding to the μ_{t_1, \dots, t_k} . Taking $W_t = 0$ for $t = 0$ shows that there exists on some (Ω, \mathcal{F}, P) a process $[W_t: t \geq 0]$ with the finite-dimensional distributions specified by the conditions (i), (ii), and (iii).

Continuity of Paths

If the Brownian motion process is to represent the motion of a particle, it is natural to require that the path functions $W(\cdot, \omega)$ be continuous. But Kolmogorov's theorem does not guarantee continuity. Indeed, for $T = [0, \infty)$, the space (Ω, \mathcal{F}) in the proof of Kolmogorov's theorem is (R^T, \mathcal{R}^T) , and as shown in the last section, the set of continuous functions does not lie in \mathcal{R}^T .

A special construction gets around this difficulty. The idea is to use for dyadic rational t the random variables W_t as already defined and then to redefine the other W_t in such a way as to ensure continuity. To carry this through requires proving that with probability 1 the sample path is uniformly continuous for dyadic rational arguments in bounded intervals.

Fix a space (Ω, \mathcal{F}, P) and on it a process $[W_t; t \geq 0]$ having the finite-dimensional distributions prescribed for Brownian motion. Let D be the set of nonnegative dyadic rationals, let $I_{nk} = [k2^{-n}, (k+2)2^{-n}]$, and put

$$(37.8) \quad \begin{aligned} M_{nk}(\omega) &= \sup_{r \in I_{nk} \cap D} |W(r, \omega) - W(k2^{-n}, \omega)| \\ M_n(\omega) &= \max_{0 \leq k < n2^n} M_{nk}(\omega). \end{aligned}$$

Suppose it is shown that $\sum P[M_n > n^{-1}]$ converges. The first Borel–Cantelli lemma will then imply that $B = [M_n > n^{-1} \text{ i.o.}]$ has probability 0. But suppose ω lies outside B . Then for every t and ϵ there exists an n such that $t < n$, $2n^{-1} < \epsilon$, and $M_n(\omega) \leq n^{-1}$. Take $\delta = 2^{-n}$. Suppose that r and r' are dyadic rationals in $[0, t]$ and $|r - r'| < \delta$. Then r and r' must for some $k < n2^n$ lie in a common interval I_{nk} (length 2×2^{-n}), in which case $|W(r, \omega) - W(r', \omega)| \leq 2M_{nk}(\omega) \leq 2M_n(\omega) \leq 2n^{-1} < \epsilon$. Therefore, $\omega \notin B$ implies that $W(r, \omega)$ is for every t uniformly continuous as r ranges over the dyadic rationals in $[0, t]$, and hence it will have a continuous extension to $[0, \infty)$.

To prove $\sum P[M_n > n^{-1}] < \infty$, use Etemadi's maximal inequality (22.10), which applies because of the independence of the increments. This, together with Markov's inequality, gives

$$\begin{aligned} P \left[\max_{i \leq 2^m} |W(t + \delta i 2^{-m}) - W(t)| > \alpha \right] \\ \leq 3 \max_{i \leq 2^m} P[|W(t + \delta i 2^{-m}) - W(t)| \geq \alpha/3] \\ \leq \frac{3}{(\alpha/3)^4} E[(W(t + \delta) - W(t))^4] = \frac{3^5}{\alpha^4} \cdot 3\delta^2 = \frac{K\delta^2}{\alpha^4} \end{aligned}$$

(see (21.7) for the moments of the normal distribution). The sets on the left here increase with m , and letting $m \rightarrow \infty$ leads to

$$(37.9) \quad P \left[\sup_{\substack{0 \leq r \leq 1 \\ r \in D}} |W(t + r\delta) - W(t)| > \alpha \right] \leq \frac{K\delta^2}{\alpha^4}.$$

Therefore,

$$P[M_n > n^{-1}] \leq n2^n \frac{K(2 \times 2^{-n})^2}{(n^{-1})^4} = \frac{4Kn^5}{2^n},$$

and $\sum P[M_n > n^{-1}]$ does converge.

Therefore, there exists a measurable set B such that $P(B) = 0$ and such that for ω outside B , $W(r, \omega)$ is uniformly continuous as r ranges over the dyadic rationals in any bounded interval. If $\omega \notin B$ and r decreases to t through dyadic rational values, then $W(r, \omega)$ has the Cauchy property and hence converges. Put

$$W'_t(\omega) = W'(t, \omega) = \begin{cases} \lim_{r \downarrow t} W(r, \omega) & \text{if } \omega \notin B, \\ 0 & \text{if } \omega \in B, \end{cases}$$

where r decreases to t through the set D of dyadic rationals. By construction, $W'(t, \omega)$ is continuous in t for each ω in Ω . If $\omega \notin B$, then $W(r, \omega) = W'(r, \omega)$ for dyadic rationals, and $W'(\cdot, \omega)$ is the continuous extension to all of $[0, \infty)$.

The next thing is to show that the W'_t have the same joint distributions as the W_t . It is convenient to prove this by a lemma which will be used again further on.

Lemma 1. *Let X_n and X be k -dimensional random vectors, and let $F_n(x)$ be the distribution function of X_n . If $X_n \rightarrow X$ with probability 1 and $F_n(x) \rightarrow F(x)$ for all x , then $F(x)$ is the distribution function of X .*

PROOF.[†] Let X have distribution function H . By two applications of Theorem 4.1, if $h > 0$, then

$$\begin{aligned} F(x_1, \dots, x_k) &= \limsup_n F_n(x_1, \dots, x_k) \leq H(x_1, \dots, x_k) \\ &\leq \liminf_n F_n(x_1 + h, \dots, x_k + h) \\ &= F(x_1 + h, \dots, x_k + h). \end{aligned}$$

It follows by continuity from above that F and H agree. ■

Now, for $0 < t_1 < \dots < t_k$, choose dyadic rationals $r_i(n)$ decreasing to the t_i . Apply Lemma 1 with $(W_{r_1(n)}, \dots, W_{r_k(n)})$ and $(W'_{t_1}, \dots, W'_{t_k})$ in the roles of X_n and X , and with the distribution function with density (37.6) in the role of F . Since (37.6) is continuous in the t_i , it follows by Scheffé's theorem that $F_n(x) \rightarrow F(x)$, and by construction $X_n \rightarrow X$ with probability 1. By the lemma, $(W'_{t_1}, \dots, W'_{t_k})$ has distribution function F , which of course is also the distribution function of $(W_{t_1}, \dots, W_{t_k})$.

Thus $[W'_t: t \geq 0]$ is a stochastic process, on the same probability space as $[W_t, t \geq 0]$, which has the finite-dimensional distributions required for Brownian motion and moreover has a continuous sample path $W'(\cdot, \omega)$ for every ω .

[†]The lemma is an obvious consequence of the weak-convergence theory of Section 29; the point of the special argument is to keep the development independent of Chapters 5 and 6.

By enlarging the set B in the definition of $W'_t(\omega)$ to include all the ω for which $W(0, \omega) \neq 0$, one can also ensure that $W'(0, \omega) = 0$. Now discard the original random variables W_t and relabel W'_t as W_t . The new $[W_t: t \geq 0]$ is a stochastic process satisfying conditions (i), (ii), and (iii) for Brownian motion and this one as well:

(iv) For each ω , $W(t, \omega)$ is continuous in t and $W(0, \omega) = 0$.

From now on, by a Brownian motion will be meant a process satisfying (iv) as well as (i), (ii), and (iii). What has been proved is this:

Theorem 37.1. *There exist processes $[W_t: t \geq 0]$ satisfying conditions (i), (ii), (iii), and (iv)—Brownian motion processes.*

In the construction above, W_r for dyadic r was used to define W_t in general. For that reason it suffices to apply Kolmogorov's theorem for a countable index set. By the second proof of that theorem, the space (Ω, \mathcal{F}, P) can be taken as the unit interval with Lebesgue measure.

The next section treats a general scheme for dealing with path-function questions by in effect replacing an uncountable time set by a countable one.

Measurable Processes

Let T be a Borel set on the line, let $[X_t: t \in T]$ be a stochastic process on an (Ω, \mathcal{F}, P) , and consider the mapping

$$(37.10) \quad (t, \omega) \rightarrow X_t(\omega) = X(t, \omega)$$

carrying $T \times \Omega$ into R^1 . Let \mathcal{T} be the σ -field of Borel subsets of T . The process is said to be *measurable* if the mapping (37.10) is measurable $\mathcal{T} \times \mathcal{F} / \mathcal{R}^1$.

In the presence of measurability, each sample path $X(\cdot, \omega)$ is measurable \mathcal{T} by Theorem 18.1. Then, for example, $\int_a^b \varphi(X(t, \omega)) dt$ makes sense if $(a, b) \subset T$ and φ is a Borel function, and by Fubini's theorem

$$E \left[\int_a^b \varphi(X(t, \cdot)) dt \right] = \int_a^b E[\varphi(X_t)] dt \quad \text{if } \int_a^b E[|\varphi(X_t)|] dt < \infty.$$

Hence the usefulness of this result:

Theorem 37.2. *Brownian motion is measurable.*

PROOF. If

$$W^{(n)}(t, \omega) = W(k2^{-n}, \omega) \quad \text{for} \quad k2^{-n} \leq t < (k+1)2^{-n}, \\ k = 0, 1, 2, \dots,$$

then the mapping $(t, \omega) \rightarrow W^{(n)}(t, \omega)$ is measurable $\mathcal{T} \times \mathcal{F}$. But by the continuity of the sample paths, this mapping converges to the mapping (37.10) pointwise (for every (t, ω)), and so by Theorem 13.4(ii) the latter mapping is also measurable $\mathcal{T} \times \mathcal{F}/\mathcal{R}^1$. ■

Irregularity of Brownian Motion Paths

Starting with a Brownian motion $[W_t: t \geq 0]$ define

$$(37.11) \quad W'_t(\omega) = c^{-1}W_{c^2t}(\omega),$$

where $c > 0$. Since $t \rightarrow c^2t$ is an increasing function, it is easy to see that the process $[W'_t: t \geq 0]$ has independent increments. Moreover, $W'_t - W'_s = c^{-1}(W_{c^2t} - W_{c^2s})$, and for $s \leq t$ this is normally distributed with mean 0 and variance $c^{-2}(c^2t - c^2s) = t - s$. Since the paths $W'(\cdot, \omega)$ all start from 0 and are continuous, $[W'_t: t \geq 0]$ is another Brownian motion. In (37.11) the time scale is contracted by the factor c^2 , but the other scale only by the factor c .

That the transformation (37.11) preserves the properties of Brownian motion implies that the paths, although continuous, must be highly irregular. It seems intuitively clear that for c large enough the path $W(\cdot, \omega)$ must with probability nearly 1 have somewhere in the time interval $[0, c]$ a chord with slope exceeding, say, 1. But then $W'(\cdot, \omega)$ has in $[0, c^{-1}]$ a chord with slope exceeding c . Since the W'_t are distributed as the W_t , this makes it plausible that $W(\cdot, \omega)$ must in arbitrarily small intervals $[0, \delta]$ have chords with arbitrarily great slopes, which in turn makes it plausible that $W(\cdot, \omega)$ cannot be differentiable at 0. More generally, mild irregularities in the path will become ever more extreme under the transformation (37.11) with ever larger values of c . It is shown below that, in fact, the paths are with probability 1 nowhere differentiable.

Also interesting in this connection is the transformation

$$(37.12) \quad W''_t(\omega) = \begin{cases} tW_{1/t}(\omega) & \text{if } t > 0, \\ 0 & \text{if } t = 0. \end{cases}$$

Again it is easily checked that the increments are independent and normally distributed with the means and variances appropriate to Brownian motion. Moreover, the path $W''(\cdot, \omega)$ is continuous except possibly at $t = 0$. But (37.9) holds with W''_s in place of W_s because it depends only on the finite-dimensional distributions, and by the continuity of $W''(\cdot, \omega)$ over $(0, \infty)$ the supremum is the same if not restricted to dyadic rationals. Therefore, $P[\sup_{s \leq n^{-3}} |W''_s| > n^{-1}] \leq K/n^2$, and it follows by the first Borel–Cantelli lemma that $W''(\cdot, \omega)$ is continuous also at 0 for ω outside a set M of probability 0. For $\omega \in M$, redefine $W''(t, \omega) \equiv 0$; then $[W''_t: t \geq 0]$ is a Brownian motion and (37.12) holds with probability 1.

The behavior of $W(\cdot, \omega)$ near 0 can be studied through the behavior of $W''(\cdot, \omega)$ near ∞ and vice versa. Since $(W_t'' - W_0'')/t = W_{1/t}$, $W''(\cdot, \omega)$ cannot have a derivative at 0 if $W(\cdot, \omega)$ has no limit at ∞ . Now, in fact,

$$(37.13) \quad \inf_n W_n = -\infty, \quad \sup_n W_n = +\infty$$

with probability 1. To prove this, note that $W_n = X_1 + \cdots + X_n$, where the $X_k = W_k - W_{k-1}$ are independent. Consider

$$\left[\sup_n W_n < \infty \right] = \bigcup_{u=1}^{\infty} \bigcap_{m=1}^{\infty} \left[\max_{i \leq m} W_i \leq u \right];$$

this is a tail set and hence by the zero-one law has probability 0 or 1. Now $-X_1, -X_2, \dots$ have the same joint distributions as X_1, X_2, \dots , and so this event has the same probability as

$$\left[\inf_n W_n > -\infty \right] = \bigcup_{u=1}^{\infty} \bigcap_{m=1}^{\infty} \left[\max_{i \leq m} (-W_i) \leq u \right].$$

If these two sets have probability 1, so has $[\sup_n |W_n| < \infty]$, so that $P[\sup_n |W_n| < x] > 0$ for some x . But $P[|W_n| < x] = P[|W_1| < x/n^{1/2}] \rightarrow 0$. This proves (37.13).

Since (37.13) holds with probability 1, $W''(\cdot, \omega)$ has with probability 1 upper and lower right derivatives of $+\infty$ and $-\infty$ at $t = 0$. The same must be true of every Brownian motion. A similar argument shows that, for each fixed t , $W(\cdot, \omega)$ is nondifferentiable at t with probability 1. In fact, $W(\cdot, \omega)$ is nowhere differentiable:

Theorem 37.3. *For ω outside a set of probability 0, $W(\cdot, \omega)$ is nowhere differentiable.*

PROOF. The proof is direct—makes no use of the transformations (37.11) and (37.12). Let

$$(37.14) \quad X_{nk} = \max \left\{ \left| W\left(\frac{k+1}{2^n}\right) - W\left(\frac{k}{2^n}\right) \right|, \left| W\left(\frac{k+2}{2^n}\right) - W\left(\frac{k+1}{2^n}\right) \right|, \right. \\ \left. \left| W\left(\frac{k+3}{2^n}\right) - W\left(\frac{k+2}{2^n}\right) \right| \right\}.$$

By independence and the fact that the differences here have the distribution of $2^{-n/2}W_1$, $P[X_{nk} \leq \epsilon] = P^3[|W_1| \leq 2^{n/2}\epsilon]$; since the standard normal density is bounded by 1, $P[X_{nk} \leq \epsilon] \leq (2 \times 2^{n/2}\epsilon)^3$. If $Y_n = \min_{k \leq n2^n} X_{nk}$, then

$$(37.15) \quad P[Y_n \leq \epsilon] \leq n2^n(2 \times 2^{n/2}\epsilon)^3.$$

Consider now the upper and lower right-hand derivatives

$$D^W(t, \omega) = \limsup_{h \downarrow 0} \frac{W(t+h, \omega) - W(t, \omega)}{h},$$

$$D_W(t, \omega) = \liminf_{h \downarrow 0} \frac{W(t+h, \omega) - W(t, \omega)}{h}.$$

Define E (not necessarily in \mathcal{F}) as the set of ω such that $D^W(t, \omega)$ and $D_W(t, \omega)$ are both finite for some value of t . Suppose that ω lies in E , and suppose specifically that

$$-K < D_W(t, \omega) \leq D^W(t, \omega) < K.$$

There exists a positive δ (depending on ω , t , and K) such that $t \leq s \leq t + \delta$ implies $|W(s, \omega) - W(t, \omega)| \leq K|s - t|$. If n exceeds some n_0 (depending on δ , K , and t), then

$$4 \times 2^{-n} < \delta, \quad 8K < n, \quad n > t.$$

Given such an n , choose k so that $(k-1)2^{-n} \leq t < k2^{-n}$. Then $|i2^{-n} - t| < \delta$ for $i = k, k+1, k+2, k+3$, and therefore $X_{nk}(\omega) \leq 2K(4 \times 2^{-n}) < n2^{-n}$. Since $k-1 \leq t2^n < n2^n$, $Y_n(\omega) \leq n2^{-n}$.

What has been shown is that if ω lies in E , then ω lies in $A_n = [Y_n \leq n2^{-n}]$ for all sufficiently large n : $E \subset \liminf_n A_n$. By (37.15),

$$P(A_n) \leq n2^n(2 \times 2^{n/2} \times n2^{-n})^3 \rightarrow 0.$$

By Theorem 4.1, $\liminf_n A_n$ has probability 0, and outside this set $W(\cdot, \omega)$ is nowhere differentiable—in fact, nowhere does it have finite upper and lower right-hand derivatives. (Similarly, outside a set of probability 0, nowhere does $W(\cdot, \omega)$ have finite upper and lower left-hand derivatives.) ■

If A is the set of ω for which $W(\cdot, \omega)$ has a derivative somewhere, what has been shown is that $A \subset B$ for a measurable B such that $P(B) = 0$; $P(A) = 0$ if A is measurable, but this has not been proved. To avoid such problems in the study of continuous-time processes, it is convenient to work in a *complete* probability space. The space (Ω, \mathcal{F}, P) is complete (see p. 44) if $A \subset B$, $B \in \mathcal{F}$, and $P(B) = 0$ together imply that $A \in \mathcal{F}$ (and then, of course, $P(A) = 0$). If the space is not already complete, it is possible to enlarge \mathcal{F} to a new σ -field and extend P to it in such a way that the new space is complete. The following assumption therefore entails no loss of generality: *For the rest of this section the space (Ω, \mathcal{F}, P) on which the Brownian motion is defined is assumed complete.* Theorem 37.3 now becomes: $W(\cdot, \omega)$ is with probability 1 nowhere differentiable.

A nowhere-differentiable path represents the motion of a particle that at no time has a velocity. Since a function of bounded variation is differentiable almost everywhere (Section 31), $W(\cdot, \omega)$ is of unbounded variation with probability 1. Such a path represents the motion of a particle that in its wanderings back and forth travels an infinite distance in finite time. The Brownian motion model thus does not in its fine structure represent physical reality. The irregularity of the Brownian motion paths is of considerable mathematical interest, however. A continuous, nowhere-differentiable function is regarded as pathological, or used to be, but from the Brownian-motion point of view such functions are the rule not the exception.[†]

The set of zeros of the Brownian motion is also interesting. By property (iv), $t = 0$ is a zero of $W(\cdot, \omega)$ for each ω . Now $[W_t'': t \geq 0]$ as defined by (37.12) is another Brownian motion, and so by (37.13) the sequence $\{W_n'': n = 1, 2, \dots\} = \{nW_{1/n}: n = 1, 2, \dots\}$ has supremum $+\infty$ and infimum $-\infty$ for ω outside a set of probability 0; for such an ω , $W(\cdot, \omega)$ changes sign infinitely often near 0 and hence by continuity has zeros arbitrarily near 0. Let $Z(\omega)$ denote the set of zeros of $W(\cdot, \omega)$. What has just been shown is that $0 \in Z(\omega)$ for each ω and that 0 is with probability 1 a limit of positive points in $Z(\omega)$. From (37.13) it also follows that $Z(\omega)$ is with probability 1 unbounded above. More is true:

Theorem 37.4. *The set $Z(\omega)$ is with probability 1 perfect [A15], unbounded, nowhere dense, and of Lebesgue measure 0.*

PROOF. Since $W(\cdot, \omega)$ is continuous, $Z(\omega)$ is closed for every ω . Let λ denote Lebesgue measure. Since Brownian motion is measurable (Theorem 37.2), Fubini's theorem applies:

$$\begin{aligned} \int_{\Omega} \lambda(Z(\omega)) P(d\omega) &= (\lambda \times P)[(t, \omega): W(t, \omega) = 0] \\ &= \int_0^{\infty} P[\omega: W(t, \omega) = 0] dt = 0. \end{aligned}$$

Thus $\lambda(Z(\omega)) = 0$ with probability 1.

If $W(\cdot, \omega)$ is nowhere differentiable, it cannot vanish throughout an interval I and hence must by continuity be nonzero throughout some subinterval of I . By Theorem 37.3, then, $Z(\omega)$ is with probability 1 nowhere dense.

It remains to show that each point of $Z(\omega)$ is a limit of other points of $Z(\omega)$. As observed above, this is true of the point 0 of $Z(\omega)$. For the general point of $Z(\omega)$, a stopping-time argument is required. Fix $r \geq 0$ and let

[†]For the construction of a specific example, see Problem 31.18.

$\tau(\omega) = \inf\{t: t \geq r, W(t, \omega) = 0\}$; note that this set is nonempty with probability 1 by (37.13). Thus $\tau(\omega)$ is the first zero following r . Now

$$[\omega: \tau(\omega) \leq t] = \left[\omega: \inf_{r \leq s \leq t} |W(s, \omega)| = 0 \right],$$

and by continuity the infimum here is unchanged if s is restricted to rationals. This shows that τ is a random variable and that

$$[\omega: \tau(\omega) \leq t] \in \sigma[W_u: u \leq t].$$

A nonnegative random variable with this property is a *stopping time*.

To know the value of τ is to know at most the values of W_u for $u \leq \tau$. Since the increments are independent, it therefore seems intuitively clear that the process

$$(37.16) \quad W_t^*(\omega) = W_{\tau(\omega)+t}(\omega) - W_{\tau(\omega)}(\omega) = W_{\tau(\omega)+t}(\omega), \quad t \geq 0,$$

ought itself to be a Brownian motion. This is, in fact, true by the next result, Theorem 37.5. What is proved there is that the finite-dimensional distributions of $[W_t^*: t \geq 0]$ are the right ones for Brownian motion. The other properties are obvious: $W^*(\cdot, \omega)$ is continuous and vanishes at 0 by construction, and the space on which $[W_t^*; t \geq 0]$ is defined is complete because it is the original space (Ω, \mathcal{F}, P) , assumed complete.

If $[W_t^*: t \geq 0]$ is indeed a Brownian motion, then, as observed above, for ω outside a set B_r of probability 0 there is a positive sequence $\{t_n\}$ such that $t_n \rightarrow 0$ and $W^*(t_n, \omega) = 0$. But then $W(\tau(\omega) + t_n, \omega) = 0$, so that $\tau(\omega)$, a zero of $W(\cdot, \omega)$, is the limit of other larger zeros of $W(\cdot, \omega)$. Now $\tau(\omega)$ was the first zero following r . (There is a different stopping time τ for each r , but the notation does not show this.) If B is the union of the B_r for rational r , the first point of $Z(\omega)$ following r is a limit of other, larger points of $Z(\omega)$. Suppose that $\omega \notin B$ and $t \in Z(\omega)$, where $t > 0$; it is to be shown that t is a limit of other points of $Z(\omega)$. If t is the limit of smaller points of $Z(\omega)$, there is of course nothing to prove. Otherwise, there is a rational r such that $r < t$ and $W(\cdot, \omega)$ does not vanish in $[r, t]$; but then, since $\omega \notin B_r$, t is a limit of larger points s that lie in $Z(\omega)$. This completes the proof of Theorem 37.4 under the provisional assumption that (37.16) is a Brownian motion. ■

The Strong Markov Property

Fix $t_0 \geq 0$ and put

$$(37.17) \quad W'_t = W_{t_0+t} - W_{t_0}, \quad t \geq 0.$$

It is easily checked that $[W'_t: t \geq 0]$ has the finite-dimensional distributions

appropriate to Brownian motion. As the other properties are obvious, it is in fact a Brownian motion.

Let

$$(37.18) \quad \mathcal{F}_t = \sigma[W_s: s \leq t].$$

The random variables (37.17) are independent of \mathcal{F}_{t_0} . To see this, suppose that $0 \leq s_1 \leq \dots \leq s_j \leq t_0$ and $0 \leq t_1 \leq \dots \leq t_k$. Put $u_i = t_0 + t_i$. Since the increments are independent, $(W'_{t_1}, W'_{t_2} - W'_{t_1}, \dots, W'_{t_k} - W'_{t_{k-1}}) = (W_{u_1} - W_{t_0}, W_{u_2} - W_{u_1}, \dots, W_{u_k} - W_{u_{k-1}})$ is independent of $(W_{s_1}, W_{s_2} - W_{s_1}, \dots, W_{s_j} - W_{s_{j-1}})$. But then $(W'_{t_1}, W'_{t_2}, \dots, W'_{t_k})$ is independent of $(W_{s_1}, W_{s_2}, \dots, W_{s_j})$. By Theorem 4.2, $(W'_{t_1}, \dots, W'_{t_k})$ is independent of \mathcal{F}_{t_0} . Thus

$$(37.19) \quad \begin{aligned} P\left(\left[(W'_{t_1}, \dots, W'_{t_k}) \in H\right] \cap A\right) \\ = P\left[(W'_{t_1}, \dots, W'_{t_k}) \in H\right] P(A) \\ = P\left[(W_{t_1}, \dots, W_{t_k}) \in H\right] P(A), \quad A \in \mathcal{F}_{t_0}, \end{aligned}$$

where the second equality follows because (37.17) is a Brownian motion. This holds for all H in \mathcal{R}^k .

The problem now is to prove all this when t_0 is replaced by a *stopping time* τ —a nonnegative random variable for which

$$(37.20) \quad [\omega: \tau(\omega) \leq t] \in \mathcal{F}_t, \quad t \geq 0.$$

It will be assumed that τ is finite, at least with probability 1. Since $[\tau = t] = [\tau \leq t] - \bigcup_n [\tau \leq t - n^{-1}]$, (37.20) implies that

$$(37.21) \quad [\omega: \tau(\omega) = t] \in \mathcal{F}_t, \quad t \geq 0.$$

The conditions (37.20) and (37.21) are analogous to the conditions (7.18) and (35.18), which prevent prevision on the part of the gambler.

Now \mathcal{F}_{t_0} contains the information on the past of the Brownian motion up to time t_0 , and the analogue for τ is needed. Let \mathcal{F}_τ consist of all measurable sets M for which

$$(37.22) \quad M \cap [\omega: \tau(\omega) \leq t] \in \mathcal{F}_t$$

for all t . (See (35.20) for the analogue in discrete time.) Note that \mathcal{F}_τ is a σ -field and τ is measurable \mathcal{F}_τ . Since $M \cap [\tau = t] = M \cap [\tau = t] \cap [\tau \leq t]$,

$$(37.23) \quad M \cap [\omega: \tau(\omega) = t] \in \mathcal{F}_t$$

for M in \mathcal{F}_τ . For example, $\tau = \inf\{t: W_t = 1\}$ is a stopping time and $[\inf_{s \leq \tau} W_s > -1]$ is in \mathcal{F}_τ .

Theorem 37.5. *Let τ be a stopping time, and put*

$$(37.24) \quad W_t^*(\omega) = W_{\tau(\omega)+t}(\omega) - W_{\tau(\omega)}(\omega), \quad t \geq 0.$$

Then $[W_t^: t \geq 0]$ is a Brownian motion, and it is independent of \mathcal{F}_τ —that is, $\sigma[W_t^*: t \geq 0]$ is independent of \mathcal{F}_τ :*

$$(37.25) \quad P\left(\left[(W_{t_1}^*, \dots, W_{t_k}^*) \in H\right] \cap M\right) \\ = P\left[(W_{t_1}^*, \dots, W_{t_k}^*) \in H\right] P(M) = P\left[(W_{t_1}, \dots, W_{t_k}) \in H\right] P(M)$$

for H in \mathcal{R}^k and M in \mathcal{F}_τ .

That the transformation (37.24) preserves Brownian motion is the *strong Markov property*.[†] Part of the conclusion is that the W_t^* are random variables.

PROOF. Suppose first that τ has countable range V and let t_0 be the general point of V . Since

$$[\omega: W_t^*(\omega) \in H] = \bigcup_{t_0 \in V} [\omega: W_{t_0+t}(\omega) - W_{t_0}(\omega) \in H, \tau(\omega) = t_0],$$

W_t^* is a random variable. Also,

$$P\left(\left[(W_{t_1}^*, \dots, W_{t_k}^*) \in H\right] \cap M\right) \\ = \sum_{t_0 \in V} P\left(\left[(W_{t_1}^*, \dots, W_{t_k}^*) \in H\right] \cap M \cap [\tau = t_0]\right).$$

If $M \in \mathcal{F}_\tau$, then $M \cap [\tau = t_0] \in \mathcal{F}_{t_0}$ by (37.23). Further, if $\tau = t_0$, then W_t^* coincides with W_t' as defined by (37.17). Therefore, (37.19) reduces this last sum to

$$\sum_{t_0 \in V} P\left[(W_{t_1}, \dots, W_{t_k}) \in H\right] P(M \cap [\tau = t_0]) \\ = P\left[(W_{t_1}, \dots, W_{t_k}) \in H\right] P(M).$$

This proves the first and third terms in (37.25) equal; to prove equality with the middle term, simply consider the case $M = \Omega$.

[†]Since the Brownian motion has independent increments, it is a Markov process (see Examples 33.9 and 33.10); hence the terminology.

Thus the theorem holds if τ has countable range. For the general τ , put

$$(37.26) \quad \tau_n = \begin{cases} k2^{-n} & \text{if } (k-1)2^{-n} < \tau \leq k2^{-n}, k = 1, 2, \dots \\ 0 & \text{if } \tau = 0. \end{cases}$$

If $k2^{-n} \leq t < (k+1)2^{-n}$, then $[\tau_n \leq t] = [\tau \leq k2^{-n}] \in \mathcal{F}_{k2^{-n}} \subset \mathcal{F}_t$. Thus each τ_n is a stopping time. Suppose that $M \in \mathcal{F}_\tau$ and $k2^{-n} \leq t < (k+1)2^{-n}$. Then $M \cap [\tau_n \leq t] = M \cap [\tau \leq k2^{-n}] \in \mathcal{F}_{k2^{-n}} \subset \mathcal{F}_t$. Thus $\mathcal{F}_\tau \subset \mathcal{F}_{\tau_n}$. Let $W_t^{(n)}(\omega) = W_{\tau_n(\omega)+t}(\omega) - W_{\tau_n(\omega)}(\omega)$ —that is, let $W_t^{(n)}$ be the W_t^* corresponding to the stopping time τ_n . If $M \in \mathcal{F}_\tau$ then $M \in \mathcal{F}_{\tau_n}$, and by an application of (37.25) to the discrete case already treated,

$$(37.27) \quad P\left([\left(W_{t_1}^{(n)}, \dots, W_{t_k}^{(n)}\right) \in H\right] \cap M\right) = P\left[\left(W_{t_1}, \dots, W_{t_k}\right) \in H\right] P(M).$$

But $\tau_n(\omega) \rightarrow \tau(\omega)$ for each ω , and by continuity of the sample paths, $W_t^{(n)}(\omega) \rightarrow W_t^*(\omega)$ for each ω . Condition on M and apply Lemma 1 with $(W_{t_1}^{(n)}, \dots, W_{t_k}^{(n)})$ for X_n , $(W_{t_1}^*, \dots, W_{t_k}^*)$ for X , and the distribution function of $(W_{t_1}, \dots, W_{t_k})$ for $F = F_n$. Then (37.25) follows from (37.27). ■

The τ in the proof of Theorem 37.4 is a stopping time, and so (37.16) is a Brownian motion, as required in that proof. Further applications will be given below.

If $\mathcal{F}^* = \sigma[W_t^*: t \geq 0]$, then according to (37.25) (and Theorem 4.2) the σ -fields \mathcal{F}_τ and \mathcal{F}^* are independent:

$$(37.28) \quad P(A \cap B) = P(A)P(B), \quad A \in \mathcal{F}_\tau, \quad B \in \mathcal{F}^*.$$

For fixed t define τ_n by (37.26) but with $t2^{-n}$ in place of 2^{-n} at each occurrence. Then $[W_\tau < x] \cap [\tau \leq t]$ is the limit superior of the sets $[W_{\tau_n} < x] \cap [\tau \leq t]$, each of which lies in \mathcal{F}_t . This proves that $[W_\tau < x]$ lies in \mathcal{F}_τ and hence that W_τ is measurable \mathcal{F}_τ . Since τ is measurable \mathcal{F}_τ ,

$$(37.29) \quad [(\tau, W_\tau) \in H] \in \mathcal{F}_\tau$$

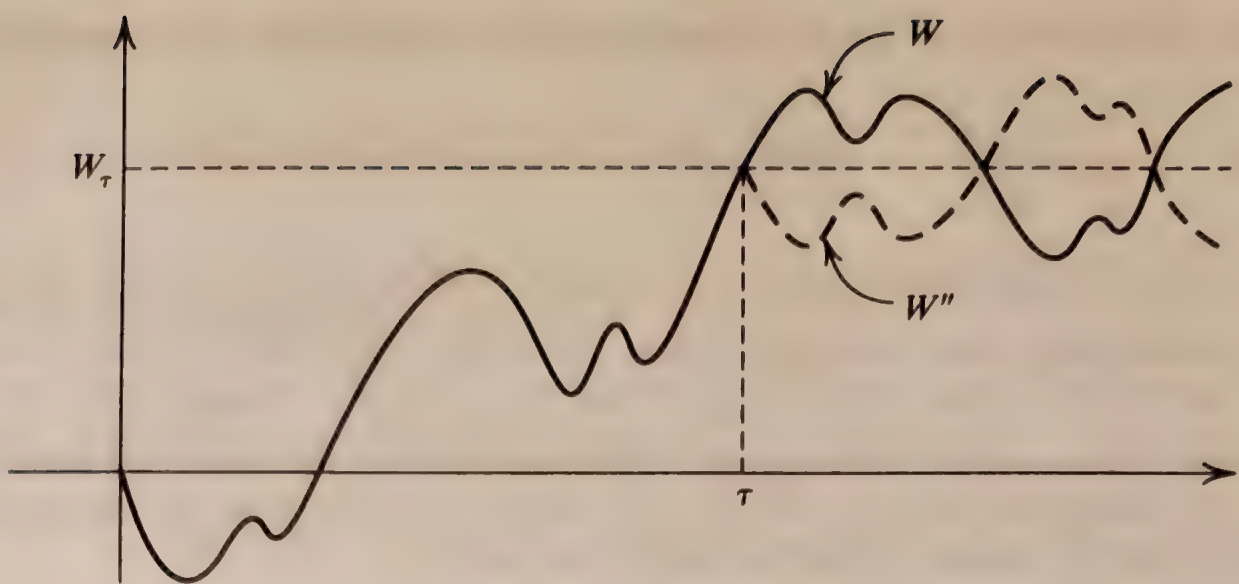
for planar Borel sets H .

The Reflection Principle

For a stopping time τ , define

$$(37.30) \quad W_t'' = \begin{cases} W_t & \text{if } t \leq \tau, \\ W_\tau - (W_t - W_\tau) & \text{if } t \geq \tau. \end{cases}$$

The sample path for $[W_t'': t \geq 0]$ is the same as the sample path for $[W_t: t \geq 0]$ up to τ , and beyond that it is reflected through the point W_τ . See the figure.



The process defined by (37.30) is a Brownian motion, and to prove it, one need only check the finite-dimensional distributions: $P[(W_{t_1}, \dots, W_{t_k}) \in H] = P[(W''_{t_1}, \dots, W''_{t_k}) \in H]$. By the argument starting with (37.26), it is enough to consider the case where τ has countable range, and for this it is enough to check the equation when the sets are intersected with $[\tau = t_0]$.

Consider for notational simplicity a pair of points:

(37.31) $P[\tau = t_0, (W_s, W_t) \in H] = P[\tau = t_0, (W''_s, W''_t) \in H].$

If $s \leq t \leq t_0$, this holds because the two events are identical. Suppose next that $s \leq t_0 \leq t$. Since $[\tau = t_0]$ lies in \mathcal{F}_{t_0} , it follows by the independence of the increments, symmetry, and the definition (37.30) that

$$\begin{aligned} P[\tau = t_0, (W_s, W_{t_0}) \in I, W_t - W_{t_0} \in J] \\ &= P[\tau = t_0, (W_s, W_{t_0}) \in I, -(W_t - W_{t_0}) \in J] \\ &= P[\tau = t_0, (W''_s, W''_{t_0}) \in I, W''_t - W''_{t_0} \in J]. \end{aligned}$$

If $K = I \times J$, this is

$$P[\tau = t_0, (W_s, W_{t_0}, W_t - W_{t_0}) \in K] = P[\tau = t_0, (W''_s, W''_{t_0}, W''_t - W''_{t_0}) \in K],$$

and by π - λ it follows for all $K \in \mathcal{R}^3$. For the appropriate K , this gives (37.31). The remaining case, $t_0 \leq s \leq t$, is similar.

These ideas can be used to derive in a very simple way the distribution of $M_t = \sup_{s \leq t} W_s$. Suppose that $x > 0$. Let $\tau = \inf[s: W_s \geq x]$, define W'' by (37.30), and put $\tau'' = \inf[s: W''_s \geq x]$ and $M''_t = \sup_{s \leq t} W''_s$. Since $\tau'' = \tau$ and W'' is another Brownian motion, reflection through the point $W_\tau = x$ shows

that

$$\begin{aligned}
 P[M_t \geq x] &= P[\tau \leq t] \\
 &= P[\tau \leq t, W_t \leq x] + P[\tau \leq t, W_t \geq x] \\
 &= P[\tau'' \leq t, W_t'' \leq x] + P[\tau \leq t, W_t \geq x] \\
 &= P[\tau'' \leq t, W_t \geq x] + P[\tau \leq t, W_t \geq x] \\
 &= P[\tau \leq t, W_t \geq x] + P[\tau \leq t, W_t \geq x] = 2P[W_t \geq x].
 \end{aligned}$$

Therefore,

$$(37.32) \quad P[M_t \geq x] = \frac{2}{\sqrt{2\pi}} \int_{x/\sqrt{t}}^{\infty} e^{-u^2/2} du.$$

This argument, an application of the *reflection principle*,[†] becomes quite transparent when referred to the diagram.

Skorohod Embedding*

Suppose that X_1, X_2, \dots are independent and identically distributed random variables with mean 0 and variance σ^2 . A powerful method, due to Skorohod, of studying the partial sums $S_n = X_1 + \dots + X_n$ is to construct an increasing sequence $\tau_0 = 0, \tau_1, \tau_2, \dots$ of stopping times such that $W(\tau_n)$ has the same distribution as S_n . The differences $\tau_k - \tau_{k-1}$ will turn out to be independent and identically distributed with mean σ^2 , so that by the law of large numbers $n^{-1}\tau_n = n^{-1}\sum_{k=1}^n(\tau_k - \tau_{k-1})$ is likely to be near σ^2 . But if τ_n is near $n\sigma^2$, then by the continuity of Brownian motion paths $W(\tau_n)$ will be near $W(n\sigma^2)$, and so the distribution of $S_n/\sigma\sqrt{n}$, which coincides with the distribution of $W(\tau_n)/\sigma\sqrt{n}$, will be near the distribution of $W(n\sigma^2)/\sigma\sqrt{n}$ —that is, will be near the standard normal distribution. The method will thus yield another proof of the central limit theorem, one independent of the characteristic-function arguments of Section 27.

But it will also give more. For example, the distribution of $\max_{k \leq n} S_k/\sigma\sqrt{n}$ is exactly the distribution of $\max_{k \leq n} W(\tau_k)/\sigma\sqrt{n}$, and this in turn is near the distribution of $\sup_{t \leq n\sigma^2} W(t)/\sigma\sqrt{n}$, which can be written down explicitly because of (37.32). It will thus be possible to derive the limiting distribution of $\max_{k \leq n} S_k$. The joint behavior of the partial sums is closely related to the behavior of Brownian motion paths.

The Skorohod construction involves the class \mathcal{T} of stopping times for which

$$(37.33) \quad E[W_\tau] = 0,$$

$$(37.34) \quad E[\tau] = E[W_\tau^2],$$

[†]See Problem 37.18 for another application.

*The rest of this section, which requires martingale theory, may be omitted.

and

$$(37.35) \quad E[\tau^2] \leq 4E[W_\tau^4].$$

Lemma 2. *All bounded stopping times are members of \mathcal{T} .*

PROOF. Define $Y_{\theta,t} = \exp(\theta W_t - \frac{1}{2}\theta^2 t)$ for all θ and for $t \geq 0$. Suppose that $s \leq t$ and $A \in \mathcal{F}_s$. Since Brownian motion has independent increments,

$$\int_A Y_{\theta,t} dP = \int_A e^{\theta W_s - \theta^2 s/2} dP \cdot E\left[e^{\theta(W_t - W_s) - \theta^2(t-s)/2}\right],$$

and a calculation with moment generating functions (see Example 21.2) shows that

$$(37.36) \quad \int_A Y_{\theta,s} dP = \int_A Y_{\theta,t} dP, \quad s \leq t, \quad A \in \mathcal{F}_s.$$

This says that for θ fixed, $[Y_{\theta,t}; t \geq 0]$ is a continuous-time martingale adapted to the σ -fields \mathcal{F}_t . It is the *moment-generating-function martingale* associated with the Brownian motion.

Let $f(\theta, t)$ denote the right side of (37.36). By Theorem 16.8,

$$\begin{aligned} \frac{\partial}{\partial \theta} f(\theta, t) &= \int_A Y_{\theta,t} (W_t - \theta t) dP, \\ \frac{\partial^2}{\partial \theta^2} f(\theta, t) &= \int_A Y_{\theta,t} [(W_t - \theta t)^2 - t] dP, \\ \frac{\partial^4}{\partial \theta^4} f(\theta, t) &= \int_A Y_{\theta,t} [(W_t - \theta t)^4 - 6(W_t - \theta t)^2 t + 3t^2] dP. \end{aligned}$$

Differentiate the other side of the equation (37.36) the same way and set $\theta = 0$. The result is

$$\int_A W_s dP = \int_A W_t dP, \quad s \leq t, \quad A \in \mathcal{F}_s,$$

$$\int_A (W_s^2 - s) dP = \int_A (W_t^2 - t) dP, \quad s \leq t, \quad A \in \mathcal{F}_s,$$

$$\int_A (W_s^4 - 6W_s^2 s + 3s^2) dP = \int_A (W_t^4 - 6W_t^2 t + 3t^2) dP, \quad s \leq t, \quad A \in \mathcal{F}_s,$$

This gives three more martingales: If Z_t is any of the three random variables

$$(37.37) \quad W_t, \quad W_t^2 - t, \quad W_t^4 - 6W_t^2 t + 3t^2,$$

then $Z_0 = 0$, Z_t is integrable and measurable \mathcal{F}_t , and

$$(37.38) \quad \int_A Z_s dP = \int_A Z_t dP, \quad s \leq t, \quad A \in \mathcal{F}_s.$$

In particular, $E[Z_t] = [Z_0] = 0$.

If τ is a stopping time with finite range $\{t_1, \dots, t_m\}$ bounded by t , then (37.38) implies that

$$E[Z_\tau] = \sum_i \int_{[\tau=t_i]} Z_{t_i} dP = \sum_i \int_{[\tau=t_i]} Z_t dP = E[Z_t] = 0.$$

Suppose that τ is bounded by t but does not necessarily have finite range. Put $\tau_n = k2^{-n}t$ if $(k-1)2^{-n}t < \tau \leq k2^{-n}t$, $1 \leq k \leq 2^n$, and put $\tau_n = 0$ if $\tau = 0$. Then τ_n is a stopping time and $E[Z_{\tau_n}] = 0$. For each of the three possibilities (37.37) for Z_t , $\sup_{s \leq t} |Z_s|$ is integrable because of (37.32). It therefore follows by the dominated convergence theorem that $E[Z_\tau] = \lim_n E[Z_{\tau_n}] = 0$.

Thus $E[Z_\tau] = 0$ for every bounded stopping time τ . The three cases (37.37) give

$$E[W_\tau] = E[W_\tau^2 - \tau] = E[W_\tau^4 - 6W_\tau^2\tau + 3\tau^2] = 0.$$

This implies (37.33), (37.34), and

$$\begin{aligned} 0 &= E[W_\tau^4] - 6E[W_\tau^2\tau] + 3E[\tau^2] \\ &\geq E[W_\tau^4] - 6E^{1/2}[W_\tau^4]E^{1/2}[\tau^2] + 3E[\tau^2]. \end{aligned}$$

If $C = E^{1/2}[W_\tau^4]$ and $x = E^{1/2}[\tau^2]$, the inequality is $0 \geq q(x) = 3x^2 - 6Cx + C^2$. Each zero of q is at most $2C$, and q is negative only between these two zeros. Therefore, $x \leq 2C$, which implies (37.35). ■

Lemma 3. Suppose that τ and τ_n are stopping times, that each τ_n is a member of \mathcal{T} , and that $\tau_n \rightarrow \tau$ with probability 1. Then τ is a member of \mathcal{T} if (i) $E[W_{\tau_n}^4] \leq E[W_\tau^4] < \infty$ for all n , or if (ii) the $W_{\tau_n}^4$ are uniformly integrable.

PROOF. Since Brownian motion paths are continuous, $W_{\tau_n} \rightarrow W_\tau$ with probability 1. Each of the two hypotheses (i) and (ii) implies that $E[W_{\tau_n}^4]$ is bounded and hence that $E[\tau_n^2]$ is bounded, and it follows (see (16.28)) that the sequences $\{\tau_n\}$, $\{W_{\tau_n}\}$, and $\{W_{\tau_n}^2\}$ are uniformly integrable. Hence (37.33) and (37.34) for τ follow by Theorem 16.14 from the same relations for the τ_n . The first hypothesis implies that $\liminf_n E[W_{\tau_n}^4] \leq E[W_\tau^4]$, and the second implies that $\lim_n E[W_{\tau_n}^4] = E[W_\tau^4]$. In either case it follows by Fatou's lemma that $E[\tau^2] \leq \liminf_n E[\tau_n^2] \leq 4 \liminf_n E[W_{\tau_n}^4] \leq 4E[W_\tau^4]$. ■

Suppose that $a, b \geq 0$ and $a + b > 0$, and let $\tau(a, b)$ be the *hitting time* for the set $\{-a, b\}$: $\tau(a, b) = \inf\{t: W_t \in \{-a, b\}\}$. By (37.13), $\tau(a, b)$ is finite with probability 1, and it is a stopping time because $\tau(a, b) \leq t$ if and only if for every m there is a rational $r \leq t$ for which W_r is within m^{-1} of $-a$ or of b . From $|W(\min\{\tau(a, b), n\})| \leq \max\{a, b\}$ it follows by Lemma 3(ii) that $\tau(a, b)$ is a member of \mathcal{T} . Since $W_{\tau(a, b)}$ assumes only the values $-a$ and b , $E[W_{\tau(a, b)}] = 0$ implies that

$$(37.39) \quad P[W_{\tau(a, b)} = -a] = \frac{b}{a+b}, \quad P[W_{\tau(a, b)} = b] = \frac{a}{a+b}.$$

This is obvious on grounds of symmetry in the case $a = b$.

Let μ be a probability measure on the line with mean 0. The program is to construct a stopping time τ for which W_τ has distribution μ . Assume that $\mu\{0\} < 1$, since otherwise $\tau \equiv 0$ obviously works. If μ consists of two point masses, they must for some positive a and b be a mass of $b/(a+b)$ at $-a$ and a mass of $a/(a+b)$ at b ; in this case $\tau_{(a, b)}$ is by (37.39) the required stopping time. The general case will be treated by adding together stopping times of this sort.

Consider a random variable X having distribution μ . (The probability space for X has nothing to do with the space the given Brownian motion is defined on.) The technique will be to represent X as the limit of a martingale X_1, X_2, \dots of a simple form and then to duplicate the martingale by $W_{\tau_1}, W_{\tau_2}, \dots$ for stopping times τ_n ; the τ_n will have a limit τ such that W_τ has the same distribution as X .

The first step is to construct sets

$$\Delta_n: a_0^{(n)} < a_1^{(n)} < \dots < a_{r_n}^{(n)}$$

and corresponding partitions

$$\mathcal{P}_n: \begin{cases} I_0^n = (-\infty, a_0^{(n)}], \\ I_k^n = (a_{k-1}^{(n)}, a_k^{(n)}], 1 \leq k \leq r_n, \\ I_{r_n+1}^n = (a_{r_n}^{(n)}, \infty). \end{cases}$$

Let $M(H)$ be the conditional mean:

$$M(H) = \frac{1}{\mu(H)} \int_H x \mu(dx) \quad \text{if } \mu(H) > 0.$$

Let Δ_1 consist of the single point $M(R^1) = E[X] = 0$, so that \mathcal{P}_1 consists of $I_0^1 = (-\infty, 0]$ and $I_1^1 = (0, \infty)$. Suppose that Δ_n and \mathcal{P}_n are given. If $\mu((I_k^n)^\circ) > 0$, split I_k^n by adding to Δ_n the point $M(I_k^n)$, which lies in $(I_k^n)^\circ$; if $\mu((I_k^n)^\circ) = 0$, I_k^n appears again in \mathcal{P}_{n+1} .

Let \mathcal{G}_n be the σ -field generated by the sets $[X \in I_k^n]$, and put $X_n = E[X | \mathcal{G}_n]$. Then X_1, X_2, \dots is a martingale and $X_n = M(I_k^n)$ on $[X \in I_k^n]$. The X_n have finite range, and their joint distributions can be written out explicitly. In fact, $[X_1 = M(I_{k_1}^1), \dots, X_n = M(I_{k_n}^n)] = [X \in I_{k_1}^1, \dots, X \in I_{k_n}^n]$, and this set is empty unless $I_{k_1}^1 \supset \dots \supset I_{k_n}^n$, in which case it is $[X_n = M(I_{k_n}^n)] = [X \in I_{k_n}^n]$. Therefore, if $k_{n-1} = j$ and $I_j^{n-1} = I_{k-1}^n \cup I_k^n$,

$$P[X_n = M(I_{k-1}^n) | X_1 = M(I_{k_1}^1), \dots, X_{n-1} = M(I_{k_{n-1}}^{n-1})] = \frac{\mu(I_{k-1}^n)}{\mu(I_j^{n-1})}$$

and

$$P[X_n = M(I_k^n) | X_1 = M(I_{k_1}^1), \dots, X_{n-1} = M(I_{k_{n-1}}^{n-1})] = \frac{\mu(I_k^n)}{\mu(I_j^{n-1})},$$

provided the conditioning event has positive probability. Thus the martingale $\{X_n\}$ has the Markov property, and if $x = M(I_j^{n-1})$, $u = M(I_{k-1}^n)$, and $v = M(I_k^n)$, then the conditional distribution of X_n given $X_{n-1} = x$ is concentrated at the two points u and v and has mean x . The structure of $\{X_n\}$ is determined by these conditional probabilities together with the distribution

$$P[X_1 = M(I_0^1)] = \mu(I_0^1), \quad P[X_1 = M(I_1^1)] = \mu(I_1^1).$$

of X_1 .

If $\mathcal{G} = \sigma(\cup_n \mathcal{G}_n)$, then $X_n \rightarrow E[X | \mathcal{G}]$ with probability 1 by the martingale theorem (Theorem 35.6). But, in fact, $X_n \rightarrow X$ with probability 1, as the following argument shows. Let B be the union of all open sets of μ -measure 0. Then B is a countable disjoint union of open intervals; enlarge B by adding to it any endpoints of μ -measure 0 these intervals may have. Then $\mu(B) = 0$, and $x \notin B$ implies that $\mu(x - \epsilon, x] > 0$ and $\mu[x, x + \epsilon) > 0$ for all positive ϵ . Suppose that $x = X(\omega) \notin B$ and let $x_n = X_n(\omega)$. Let $I_{k_n}^n$ be the element of \mathcal{G}_n containing x ; then $x_{n+1} = M(I_{k_n}^n)$ and $I_{k_n}^n \downarrow I$ for some interval I . Suppose that $x_{n+1} < x - \epsilon$ for n in an infinite sequence N of integers. Then x_{n+1} is the left endpoint of $I_{k_{n+1}}^{n+1}$ for n in N and converges along N to the left endpoint, say a , of I , and $(x - \epsilon, x] \subset I$. Further, $x_{n+1} = M(I_{k_n}^n) \rightarrow M(I)$ along N , so that $M(I) = a$. But this is impossible because $\mu(x - \epsilon, x] > 0$. Therefore, $x_n \geq x - \epsilon$ for large n . Similarly, $x_n \leq x + \epsilon$ for large n , and so $x_n \rightarrow x$. Thus $X_n(\omega) \rightarrow X(\omega)$ if $X(\omega) \notin B$, the probability of which is 1.

Now $X_1 = E[X | \mathcal{G}_1]$ has mean 0, and its distribution consists of point masses at $-a = M(I_0^1)$ and $b = M(I_1^1)$. If $\tau_1 = \tau(a, b)$ is the hitting time to

$\{-a, b\}$, then (see (37.39)) τ_1 is a stopping time, a member of \mathcal{T} , and W_{τ_1} has the same distribution as X_1 .

Let τ_2 be the infimum of those t for which $t \geq \tau_1$ and W_t is one of the points $M(I_k^2)$, $0 \leq k \leq r_2 + 1$. By (37.13), τ_2 is finite with probability 1; it is a stopping time, because $\tau_2 \leq t$ if and only if for every m there are rationals r and s such that $r \leq s + m^{-1}$, $r \leq t$, $s \leq t$, W_r is within m^{-1} of one of the points $M(I_j^1)$, and W_s is within m^{-1} of one of the points $M(I_k^2)$. Since $|W(\min\{\tau_2, n\})|$ is at most the maximum of the values $|M(I_k^2)|$, it follows by Lemma 3(ii) that τ_2 is a member of \mathcal{T} .

Define W_t^* by (37.24) with τ_1 for τ . If $x = M(I_j^1)$, then x is an endpoint common to two adjacent intervals I_{k-1}^2 and I_k^2 ; put $u = M(I_{k-1}^2)$ and $v = M(I_k^2)$. If $W_{\tau_1} = x$, then u and v are the only possible values of W_{τ_2} . If τ^* is the first time the Brownian motion $[W_t^*; t \geq 0]$ hits $u - x$ or $v - x$, then by (37.39),

$$P[W_{\tau^*}^* = u - x] = \frac{v - x}{v - u}, \quad P[W_{\tau^*}^* = v - x] = \frac{x - u}{v - u}.$$

On the set $[W_{\tau_1} = x]$, τ_2 coincides with $\tau_1 + \tau^*$, and it follows by (37.28) that

$$\begin{aligned} P[W_{\tau_1} = x, W_{\tau_2} = v] &= P[W_{\tau_1} = x, x + W_{\tau^*}^* = v] \\ &= P[W_{\tau_1} = x] P[W_{\tau^*}^* = v - x] = P[W_{\tau_1}] \frac{x - u}{v - u}. \end{aligned}$$

This, together with the same computation with u in place of v , shows that for $W_{\tau_1} = x$ the conditional distribution of W_{τ_2} is concentrated at the two points u and v and has mean x . Thus the conditional distribution of W_{τ_2} given W_{τ_1} coincides with the conditional distribution of X_2 given X_1 . Since W_{τ_1} and X_1 have the same distribution, the random vectors (W_{τ_1}, W_{τ_2}) and (X_1, X_2) also have the same distribution.

An inductive extension of this argument proves the existence of a sequence of stopping times τ_n such that $\tau_1 \leq \tau_2 \leq \dots$, each τ_n is a member of \mathcal{T} , and for each n , $W_{\tau_1}, \dots, W_{\tau_n}$ have the same joint distribution as X_1, \dots, X_n . Now suppose that X has finite variance. Since τ_n is a member of \mathcal{T} , $E[\tau_n] = E[W_{\tau_n}^2] = E[X_n^2] = E[E^2(X | \mathcal{G}_n)] \leq E[X^2]$ by Jensen's inequality (34.7). Thus $\tau = \lim_n \tau_n$ is finite with probability 1. Obviously it is a stopping time, and by path continuity, $W_{\tau_n} \rightarrow W_\tau$ with probability 1. Since $X_n \rightarrow X$ with probability 1, it is a consequence of the following lemma that W_τ has the distribution of X .

Lemma 4. *If $X_n \rightarrow X$ and $Y_n \rightarrow Y$ with probability 1, and if X_n and Y_n have the same distribution, then so do X and Y .*

PROOF.[†] By two applications of (4.9),

$$\begin{aligned} P[X \leq x] &\leq P[X < x + \epsilon] \leq \liminf_n P[X_n \leq x + \epsilon] \\ &\leq \limsup_n P[Y_n \leq x + \epsilon] \leq P[Y \leq x + \epsilon]. \end{aligned}$$

Let $\epsilon \rightarrow 0$: $P[X \leq x] \leq P[Y \leq x]$. Now interchange the roles of X and Y . ■

Since $X_n^2 \leq E[X^2 | \mathcal{G}_n]$, the X_n are uniformly integrable by the lemma preceding Theorem 35.6. By the monotone convergence theorem and Theorem 16.14, $E[\tau] = \lim_n E[\tau_n] = \lim_n E[W_{\tau_n}^2] = \lim_n E[X_n^2] = E[X^2] = E[W_\tau^2]$. If $E[X^4] < \infty$, then $E[W_{\tau_n}^4] = E[X_n^4] \leq E[X^4] = E[W_\tau^4]$ (Jensen's inequality again), and so τ is a member of \mathcal{T} . Hence $E[\tau^2] \leq 4E[W_\tau^4]$.

This construction establishes the first of Skorohod's embedding theorems:

Theorem 37.6. *Suppose that X is a random variable with mean 0 and finite variance. There is a stopping time τ such that W_τ has the same distribution as X , $E[\tau] = E[X^2]$, and $E[\tau^2] \leq 4E[X^4]$.*

Of course, the last inequality is trivial unless $E[X^4]$ is finite. The theorem could be stated in terms not of X but of its distribution, the point being that the probability space X is defined on is irrelevant. Skorohod's second embedding theorem is this:

Theorem 37.7. *Suppose that X_1, X_2, \dots are independent and identically distributed random variables with mean 0 and finite variance, and put $S_n = X_1 + \dots + X_n$. There is a nondecreasing sequence τ_1, τ_2, \dots of stopping times such that the W_{τ_n} have the same joint distributions as the S_n and $\tau_1, \tau_2 - \tau_1, \tau_3 - \tau_2, \dots$ are independent and identically distributed random variables satisfying $E[\tau_n - \tau_{n-1}] = E[X_1^2]$ and $E[(\tau_n - \tau_{n-1})^2] \leq 4E[X_1^4]$.*

PROOF. The method is to repeat the construction above inductively. For notational clarity write $W_t = W_t^{(1)}$ and put $\mathcal{F}_t^{(1)} = \sigma[W_s^{(1)}: 0 \leq s \leq t]$ and $\mathcal{F}^{(1)} = \sigma[W_t^{(1)}: t \geq 0]$. Let δ_1 be the stopping time of Theorem 37.6, so that $W_{\delta_1}^{(1)}$ and X_1 have the same distribution. Let $\mathcal{F}_{\delta_1}^{(1)}$ be the class of M such that $M \cap [\delta_1 \leq t] \in \mathcal{F}_t^{(1)}$ for all t .

Now put $W_t^{(2)} = W_{\delta_1+t}^{(1)} - W_{\delta_1}^{(1)}$, $\mathcal{F}_t^{(2)} = \sigma[W_s^{(2)}: 0 \leq s \leq t]$, and $\mathcal{F}^{(2)} = \sigma[W_t^{(2)}: t \geq 0]$. By another application of Theorem 37.6, construct a stopping time δ_2 for the Brownian motion $[W_t^{(2)}: t \geq 0]$ in such a way that $W_{\delta_2}^{(2)}$ has the same distribution as X_1 . In fact, use for δ_2 the very same martingale construction as for δ_1 , so that $(\delta_1, W_{\delta_1}^{(1)})$ and $(\delta_2, W_{\delta_2}^{(2)})$ have the same distribution. Since $\mathcal{F}_{\delta_1}^{(1)}$ and $\mathcal{F}^{(2)}$ are independent (see (37.28)), it follows (see (37.29)) that $(\delta_1, W_{\delta_1}^{(1)})$ and $(\delta_2, W_{\delta_2}^{(2)})$ are independent.

[†]This is obvious from the weak-convergence point of view.

Let $\mathcal{F}_{\delta_2}^{(2)}$ be the class of M such that $M \cap [\delta_2 \leq t] \in \mathcal{F}_t^{(2)}$ for all t . If $W_t^{(3)} = W_{\delta_2+t}^{(2)} - W_{\delta_2}^{(2)}$ and $\mathcal{F}^{(3)}$ is the σ -field generated by these random variables, then again $\mathcal{F}_{\delta_2}^{(2)}$ and $\mathcal{F}^{(3)}$ are independent. These two σ -fields are contained in $\mathcal{F}^{(2)}$, which is independent of $\mathcal{F}_{\delta_1}^{(1)}$. Therefore, the three σ -fields $\mathcal{F}_{\delta_1}^{(1)}$, $\mathcal{F}_{\delta_2}^{(2)}$, $\mathcal{F}^{(3)}$ are independent. The procedure therefore extends inductively to give independent, identically distributed random vectors $(\delta_n, W_{\delta_n}^{(n)})$. If $\tau_n = \delta_1 + \cdots + \delta_n$, then $W_{\tau_n}^{(1)} = W_{\delta_1}^{(1)} + \cdots + W_{\delta_n}^{(n)}$ has the distribution of $X_1 + \cdots + X_n$. ■

Invariance*

If $E[X_1^2] = \sigma^2$, then, since the random variables $\tau_n - \tau_{n-1}$ of Theorem 37.7 are independent and identically distributed, the strong law of large numbers (Theorem 22.1) applies and hence so does the weak one:

$$(37.40) \quad P[|n^{-1}\tau_n - \sigma^2| \geq \epsilon] \rightarrow 0.$$

(If $E[X_1^4] < \infty$, so that the $\tau_n - \tau_{n-1}$ have second moments, this follows immediately by Chebyshev's inequality.) Now S_n has the distribution of $W(\tau_n)$, and τ_n is near $n\sigma^2$ by (37.40); hence S_n should have nearly the distribution of $W(n\sigma^2)$, namely the normal distribution with mean 0 and variance $n\sigma^2$.

To prove this, choose an increasing sequence of integers N_k such that $P[|n^{-1}\tau_n - \sigma^2| \geq k^{-1}] < k^{-1}$ for $n \geq N_k$, and put $\epsilon_n = k^{-1}$ for $N_k \leq n < N_{k+1}$. Then $\epsilon_n \rightarrow 0$ and $P[|n^{-1}\tau_n - \sigma^2| \geq \epsilon_n] < \epsilon_n$. By two applications of (37.32),

$$\begin{aligned} \delta_n(\epsilon) &= P\left[\frac{|W(n\sigma^2) - W(\tau_n)|}{\sigma\sqrt{n}} \geq \epsilon\right] \\ &\leq P[|n^{-1}\tau_n - \sigma^2| \geq \epsilon_n] + P\left[\sup_{|t - n\sigma^2| \leq \epsilon_n n} |W(t) - W(n\sigma^2)| \geq \epsilon\sigma\sqrt{n}\right] \\ &\leq \epsilon_n + 4P[|W(\epsilon_n n)| \geq \epsilon\sigma\sqrt{n}], \end{aligned}$$

and it follows by Chebyshev's inequality that $\lim_n \delta_n(\epsilon) = 0$. Since S_n is distributed as $W(\tau_n)$,

$$\begin{aligned} P\left[\frac{W(n\sigma^2)}{\sigma\sqrt{n}} \leq x - \epsilon\right] - \delta_n(\epsilon) &\leq P\left[\frac{S_n}{\sigma\sqrt{n}} \leq x\right] \\ &\leq P\left[\frac{W(n\sigma^2)}{\sigma\sqrt{n}} \leq x + \epsilon\right] + \delta_n(\epsilon). \end{aligned}$$

*This topic may be omitted.

Here $W(n\sigma^2)/\sigma\sqrt{n}$ can be replaced by a random variable N with the standard normal distribution, and letting $n \rightarrow \infty$ and then $\epsilon \rightarrow 0$ shows that

$$\lim_n P\left[\frac{S_n}{\sigma\sqrt{n}} \leq x\right] = P[N \leq x].$$

This gives a new proof of the central limit theorem for independent, identically distributed random variables with second moments (the Lindeberg–Lévy theorem—Theorem 27.1). Observe that none of the convergence theory of Chapter 5 has been used.

This proof of the central limit theorem is an application of the *invariance principle*: S_n has nearly the distribution of $W(n\sigma^2)$, and the distribution of the latter does not depend on (vary with) the distribution common to the X_n . More can be said if the X_n have fourth moments.

For each n , define a stochastic process $[Y_n(t): 0 \leq t \leq 1]$ by $Y_n(0, \omega) = 0$ and

$$(37.41) \quad Y_n(t, \omega) = \frac{1}{\sigma\sqrt{n}} S_k(\omega) \quad \text{if } \frac{k-1}{n} < t \leq \frac{k}{n}, \quad k = 1, \dots, n.$$

If $k/n = t > 0$ and n is large, then k is large, too, and $Y_n(t) = t^{1/2} S_k / \sigma\sqrt{k}$ is by the central limit theorem approximately normally distributed with mean 0 and variance t . Since the X_n are independent, the increments of (37.41) should be approximately independent, and so the process should behave approximately as a Brownian motion does.

Let τ_n be the stopping times of Theorem 37.7, and in analogy with (37.41) put $Z_n(0) = 0$ and

$$(37.42) \quad Z_n(t) = \frac{1}{\sigma\sqrt{n}} W(\tau_k) \quad \text{if } \frac{k-1}{n} < t \leq \frac{k}{n}, \quad k = 1, \dots, n.$$

By construction, the finite-dimensional distributions of $[Y_n(t): 0 \leq t \leq 1]$ coincide with those of $[Z_n(t): 0 \leq t \leq 1]$. It will be shown that the latter process nearly coincides with $[W(tn\sigma^2)/\sigma\sqrt{n}: 0 \leq t \leq 1]$, which is itself a Brownian motion over the time interval $[0, 1]$ —see (37.11). Put $W_n(t) = W(tn\sigma^2)/\sigma\sqrt{n}$.

Let $B_n(\delta)$ be the event that $|\tau_k - k\sigma^2| \geq \delta n\sigma^2$ for some $k \leq n$. By Kolmogorov's inequality (22.9),

$$(37.43) \quad P(B_n(\delta)) \leq \frac{\text{Var}[\tau_n]}{\delta^2 n^2 \sigma^4} \leq \frac{4E[X_1^4]}{\delta^2 n \sigma^4} \rightarrow 0.$$

If $(k-1)n^{-1} < t \leq kn^{-1}$ and $n > \delta^{-1}$, then

$$\left| \frac{\tau_k}{n\sigma^2} - t \right| \leq \left| \frac{\tau_k}{n\sigma^2} - \frac{k}{n} \right| + \frac{1}{n} \leq 2\delta$$

on the event $(B_n(\delta))^c$, and so

$$|Z_n(t) - W_n(t)| = \left| W_n\left(\frac{\tau_k}{n\sigma^2}\right) - W_n(t) \right| \leq \sup_{|s-t| \leq 2\delta} |W_n(s) - W_n(t)|$$

on $(B_n(\delta))^c$. Since the distribution of this last random variable is unchanged if the $W_n(t)$ are replaced by $W(t)$,

$$\begin{aligned} P\left[\sup_{t \leq 1} |Z_n(t) - W_n(t)| \geq \epsilon\right] \\ \leq P(B_n(\delta)) + P\left[\sup_{t \leq 1} \sup_{|s-t| \leq 2\delta} |W(s) - W(t)| \geq \epsilon\right]. \end{aligned}$$

Let $n \rightarrow \infty$ and then $\delta \rightarrow 0$; it follows by (37.43) and the continuity of Brownian motion paths that

$$(37.44) \quad \lim_n P\left[\sup_{t \leq 1} |Z_n(t) - W_n(t)| \geq \epsilon\right] = 0$$

for positive ϵ . Since the processes (37.41) and (37.42) have the same finite-dimensional distributions, this proves the following general invariance principle or *functional central limit theorem*.

Theorem 37.8. *Suppose that X_1, X_2, \dots are independent, identically distributed random variables with mean 0, variance σ^2 , and finite fourth moments, and define $Y_n(t)$ by (37.41). There exist (on another probability space), for each n , processes $[Z_n(t): 0 \leq t \leq 1]$ and $[W_n(t): 0 \leq t \leq 1]$ such that the first has the same finite-dimensional distributions as $[Y_n(t): 0 \leq t \leq 1]$, the second is a Brownian motion, and $P[\sup_{t \leq 1} |Z_n(t) - W_n(t)| \geq \epsilon] \rightarrow 0$ for positive ϵ .*

As an application, consider the maximum $M_n = \max_{k \leq n} S_k$. Now $M_n/\sigma\sqrt{n} = \sup_t Y_n(t)$ has the same distribution as $\sup_t Z_n(t)$, and it follows by (37.44) that

$$P\left[\left|\sup_{t \leq 1} Z_n(t) - \sup_{t \leq 1} W_n(t)\right| \geq \epsilon\right] \rightarrow 0.$$

But $P[\sup_{t \leq 1} W_n(t) \geq x] = P[\sup_{t \leq 1} W(t) \geq x] = 2P[N \geq x]$ for $x \geq 0$ by (37.32). Therefore,

$$(37.45) \quad P\left[\frac{M_n}{\sigma\sqrt{n}} \leq x\right] \rightarrow 2P[N \leq x], \quad x \geq 0.$$

PROBLEMS

37.1. 36.2 \uparrow Show that $K(s, t) = \min\{s, t\}$ is nonnegative-definite; use Problem 36.2 to prove the existence of a process with the finite-dimensional distributions prescribed for Brownian motion.

37.2. Let $X(t)$ be independent, standard normal variables, one for each dyadic rational t (Theorem 20.4; the unit interval can be used as the probability space). Let $W(0) = 0$ and $W(n) = \sum_{k=1}^n X(k)$. Suppose that $W(t)$ is already defined for dyadic rationals of rank n , and put

$$W\left(\frac{2k+1}{2^{n+1}}\right) = \frac{1}{2}W\left(\frac{k}{2^n}\right) + \frac{1}{2}W\left(\frac{k+1}{2^n}\right) + \frac{1}{2^{1+n/2}}X\left(\frac{2k+1}{2^{n+1}}\right).$$

Prove by induction that the $W(t)$ for dyadic t have the finite-dimensional distributions prescribed for Brownian motion. Now construct a Brownian motion with continuous paths by the argument leading to Theorem 37.1. This avoids an appeal to Kolmogorov's existence theorem.

37.3. \uparrow For each n define new variables $W_n(t)$ by setting $W_n(k/2^n) = W(k/2^n)$ for dyadics of order n and interpolating linearly in between. Set $\delta_n = \sup_{t \leq n} |W_{n+1}(t) - W_n(t)|$, and show that

$$\delta_n = \max_{0 \leq k < n2^n} \left| W\left(\frac{2k+1}{2^{n+1}}\right) - \left[\frac{1}{2}W\left(\frac{k}{2^n}\right) + \frac{1}{2}W\left(\frac{k+1}{2^n}\right) \right] \right|.$$

The construction in the preceding problem makes it clear that the difference here is normal with variance $1/2^{n+2}$. Find positive x_n such that $\sum x_n$ and $\sum P[\delta_n \geq x_n]$ both converge, and conclude that outside a set of probability 0, $W_n(t, \omega)$ converges uniformly over bounded intervals. Replace $W(t, \omega)$ by $\lim_n W_n(t, \omega)$. This gives another construction of a Brownian motion with continuous paths.

37.4. 36.6 \uparrow Let $T = [0, \infty)$, and let P be a probability measure on (R^T, \mathcal{R}^T) having the finite-dimensional distributions prescribed for Brownian motion. Let C consist of the continuous elements of R^T .

(a) Show that $P_*(C) = 0$, or $P^*(R^T - C) = 1$ (see (3.9) and (3.10)). Thus completing (R^T, \mathcal{R}^T, P) will not give C probability 1.

(b) Show that $P^*(C) = 1$.

37.5. Suppose that $[W_t: t \geq 0]$ is some stochastic process having independent, stationary increments satisfying $E[W_t] = 0$ and $E[W_t^2] = t$. Show that if the finite-dimensional distributions are preserved by the transformation (37.11), then they must be those of Brownian motion.

37.6. Show that $\bigcap_{t > 0} \sigma[W_s: s \geq t]$ contains only sets of probability 0 and 1. Do the same for $\bigcap_{\epsilon > 0} \sigma[W_t: 0 < t < \epsilon]$; give examples of sets in this σ -field.

37.7. Show by a direct argument that $W(\cdot, \omega)$ is with probability 1 of unbounded variation on $[0, 1]$: Let $Y_n = \sum_{i=1}^{2^n} |W(i2^{-n}) - W((i-1)2^{-n})|$. Show that Y_n has mean $2^{n/2}E[|W_1|]$ and variance at most $\text{Var}[|W_1|]$. Conclude that $\sum P[Y_n < n] < \infty$.

- 37.8. Show that the Poisson process as defined by (23.5) is measurable.
- 37.9. Show that for $T = [0, \infty)$ the coordinate-variable process $[Z_t: t \in T]$ on (R^T, \mathcal{R}^T) is not measurable.
- 37.10. Extend Theorem 37.4 to the set $[t: W(t, \omega) = \alpha]$.
- 37.11. Let τ_α be the first time the Brownian motion hits $\alpha > 0$: $\tau_\alpha = \inf\{t: W_t \geq \alpha\}$. Show that the distribution of τ_α has over $(0, \infty)$ the density
- $$(37.46) \quad h_\alpha(t) = \frac{\alpha}{\sqrt{2\pi}} \frac{1}{t^{3/2}} e^{-\alpha^2/2t}.$$
- Show that $E[\tau_\alpha] = \infty$. Show that τ_α has the same distribution as α^2/N^2 , where N is a standard normal variable.
- 37.12. \uparrow (a) Show by the strong Markov property that τ_α and $\tau_{\alpha+\beta} - \tau_\alpha$ are independent and that the latter has the same distribution as τ_β . Conclude that $h_\alpha * h_\beta = h_{\alpha+\beta}$. Show that $\beta\tau_\alpha$ has the same distribution as $\tau_\alpha \sqrt{\beta}$.
(b) Show that each h_α is stable—see Problem 28.10.
- 37.13. \uparrow Suppose that X_1, X_2, \dots are independent and each has the distribution (37.46).
(a) Show that $(X_1 + \dots + X_n)/n^2$ also has the distribution (37.46). Contrast this with the law of large numbers.
(b) Show that $P[n^{-2} \max_{k \leq n} X_k \leq x] \rightarrow \exp(-\alpha\sqrt{2/\pi x})$ for $x > 0$. Relate this to Theorem 14.3.
- 37.14. 37.11 \uparrow Let $\rho(s, t)$ be the probability that a Brownian path has at least one zero in (s, t) . From (37.46) and the Markov property deduce

$$(37.47) \quad \rho(s, t) = \frac{2}{\pi} \arccos \sqrt{\frac{s}{t}}.$$

Hint: Condition with respect to W_s .

- 37.15. \uparrow (a) Show that the probability of no zero in $(t, 1)$ is $(2/\pi) \arcsin \sqrt{t}$ and hence that the position of the last zero preceding 1 is distributed over $(0, 1)$ with density $\pi^{-1}(t(1-t))^{-1/2}$.
(b) Similarly calculate the distribution of the position of the first zero following time 1.
(c) Calculate the joint distribution of the two zeros in (a) and (b).
- 37.16. \uparrow (a) Show by Theorem 37.8 that $\inf_{s \leq u \leq t} Y_n(u)$ and $\inf_{s \leq u \leq t} Z_n(u)$ both converge in distribution to $\inf_{s \leq u \leq t} W(u)$ for $0 \leq s \leq t \leq 1$. Prove a similar result for the supremum.
(b) Let $A_n(s, t)$ be the event that S_k , the position at time k in a symmetric random walk, is 0 for at least one k in the range $sn \leq k \leq tn$, and show that $P(A_n(s, t)) \rightarrow (2/\pi) \arccos \sqrt{s/t}$.

(c) Let T_n be the maximum k such that $k \leq n$ and $S_k = 0$. Show that T_n/n has asymptotically the distribution with density $\pi^{-1}(t(1-t))^{-1/2}$ over $(0, 1)$. As this density is larger at the ends of the interval than in the middle, the last time during a night's play a gambler was even is more likely to be either early or late than to be around midnight.

37.17. \uparrow Show that $\rho(s, t) = \rho(t^{-1}, s^{-1}) = \rho(cs, ct)$. Check this by (37.47) and also by the fact that the transformations (37.11) and (37.12) preserve the properties of Brownian motion.

37.18. Deduce by the reflection principle that (M_t, W_t) has density

$$\frac{2(2y-x)}{t\sqrt{2\pi t}} \exp\left[-\frac{(2y-x)^2}{2t}\right]$$

on the set where $y \geq 0$ and $y \geq x$. Now deduce from Theorem 37.8 the corresponding limit theorem for symmetric random walk.

37.19. Show by means of the transformation (37.12) that for positive a and b the probability is 1 that the process is within the boundary $-at < W_t < bt$ for all sufficiently large t . Show that $a/(a+b)$ is the probability that it last touches above rather than below.

37.20. The martingale calculation used for (37.39) also works for slanting boundaries. For positive a, b, r , let τ be the smallest t such that either $W_t = -a + rt$ or $W_t = b + rt$, and let $p(a, b, r)$ be the probability that the exit is through the upper barrier—that $W_\tau = b + r\tau$.

(a) For the martingale $Y_{\theta, t}$ in the proof of Lemma 2, show that $E[Y_{\theta, t}] = 1$. Operating formally at first, conclude that

$$(37.48) \quad E[e^{\theta W_\tau - \frac{1}{2}\theta^2 \tau}] = 1.$$

Take $\theta = 2r$, and note that $\theta W_\tau - \frac{1}{2}\theta^2 \tau$ is then $2rb$ if the exit is above (probability $p(a, b, r)$) and $-2ra$ if the exit is below (probability $1 - p(a, b, r)$). Deduce

$$p(a, b, r) = \frac{1 - e^{-2ra}}{e^{2rb} - e^{-2ra}}.$$

(b) Show that $p(a, b, r) \rightarrow a/(a+b)$ as $r \rightarrow 0$, in agreement with (37.39).

(c) It remains to justify (37.48) for $\theta = 2r$. From $E[Y_{\theta, t}] = 1$ deduce

$$(37.49) \quad E[e^{2r(W_\sigma - r^2 \sigma)}] = 1$$

for nonrandom σ . By the arguments in the proofs of Lemmas 2 and 3, show that (37.49) holds for simple stopping times σ , for bounded ones, for $\sigma = \tau \wedge n$, for $\sigma = \tau$.

SECTION 38. NONDENumerable PROBABILITIES*

Introduction

As observed a number of times above, the finite-dimensional distributions do not suffice to determine the character of the sample paths of a process. To obtain paths with natural regularity properties, the Poisson and Brownian motion processes were constructed by ad hoc methods. It is always possible to ensure that the paths have a certain very general regularity property called *separability*, and from this property will follow in appropriate circumstances various other desirable regularity properties.

Section 4 dealt with “denumerable” probabilities; questions about path functions involve all the time points and hence concern “nondenumerable” probabilities.

Example 38.1. For a mathematically simple illustration of the fact that path properties are not entirely determined by the finite-dimensional distributions, consider a probability space (Ω, \mathcal{F}, P) on which is defined a positive random variable V with continuous distribution: $P[V = x] = 0$ for each x . For $t \geq 0$, put $X(t, \omega) = 0$ for all ω , and put

$$(38.1) \quad Y(t, \omega) = \begin{cases} 1 & \text{if } V(\omega) = t, \\ 0 & \text{if } V(\omega) \neq t. \end{cases}$$

Since V has continuous distribution, $P[X_t = Y_t] = 1$ for each t , and so $[X_t: t \geq 0]$ and $[Y_t: t \geq 0]$ are stochastic processes with identical finite-dimensional distributions; for each t_1, \dots, t_k , the distribution $\mu_{t_1 \dots t_k}$ common to $(X_{t_1}, \dots, X_{t_k})$ and $(Y_{t_1}, \dots, Y_{t_k})$ concentrates all its mass at the origin of R^k . But what about the sample paths? Of course, $X(\cdot, \omega)$ is identically 0, but $Y(\cdot, \omega)$ has a discontinuity—it is 1 at $t = V(\omega)$ and 0 elsewhere. It is because the position of this discontinuity has a continuous distribution that the two processes have the same finite-dimensional distributions. ■

Definitions

The idea of separability is to make a countable set of time points serve to determine the properties of the process. In all that follows, the time set T will for definiteness be taken as $[0, \infty)$. Most of the results hold with an arbitrary subset of the line in the role of T .

As in Section 36, let R^T be the set of all real functions over $T = [0, \infty)$. Let D be a countable, dense subset of T . A function x —an element of R^T —is *separable* D , or *separable with respect to* D , if for each t in T there exists a

*This section may be omitted.

sequence t_1, t_2, \dots of points such that

$$(38.2) \quad t_n \in D, \quad t_n \rightarrow t, \quad x(t_n) \rightarrow x(t).$$

(Because of the middle condition here, it was redundant to require D dense at the outset.) For t in D , (38.2) imposes no condition on x , since t_n may be taken as t . An x separable with respect to D is determined by its values at the points of D . Note, however, that separability requires that (38.2) hold for every t —an uncountable set of conditions. It is not hard to show that the set of functions separable with respect to D lies outside \mathcal{R}^T .

Example 38.2. If x is everywhere continuous or right-continuous, then it is separable with respect to every countable, dense D .

Suppose that $x(t)$ is 0 for $t \neq v$ and 1 for $t = v$, where $v > 0$. Then x is not separable with respect to D unless v lies in D . The paths $Y(\cdot, \omega)$ in Example 38.1 are of this form. ■

The condition for separability can be stated another way: x is separable D if and only if for every t and every open interval I containing t , $x(t)$ lies in the closure of $[x(s): s \in I \cap D]$.

Suppose that x is separable D and that I is an open interval in T . If $\epsilon > 0$, then $x(t_0) + \epsilon > \sup_{t \in I} x(t) = u$ for some t_0 in I . By separability $|x(s_0) - x(t_0)| < \epsilon$ for some s_0 in $I \cap D$. But then $x(s_0) + 2\epsilon > u$, so that

$$(38.3) \quad \sup_{t \in I} x(t) = \sup_{t \in I \cap D} x(t).$$

Similarly,

$$(38.4) \quad \inf_{t \in I} x(t) = \inf_{t \in I \cap D} x(t)$$

and

$$(38.5) \quad \sup_{t_0 \leq t < t_0 + \delta} |x(t) - x(t_0)| = \sup_{\substack{t_0 \leq t < t_0 + \delta \\ t \in D}} |x(t) - x(t_0)|.$$

A stochastic process $[X_t: t \geq 0]$ on (Ω, \mathcal{F}, P) is separable D if D is a countable, dense subset of $T = [0, \infty)$ and there is an \mathcal{F} -set N such that $P(N) = 0$ and such that the sample path $X(\cdot, \omega)$ is separable with respect to D for ω outside N . Finally, the process is separable if it is separable with respect to some D ; this D is sometimes called a *separant*. In these definitions it is assumed for the moment that $X(t, \omega)$ is a finite real number for each t and ω .

Example 38.3. If the sample path $X(\cdot, \omega)$ is continuous for each ω , then the process is separable with respect to each countable, dense D . This covers Brownian motion as constructed in the preceding section. ■

Example 38.4. Suppose that $[W_t: t \geq 0]$ has the finite-dimensional distributions of Brownian motion, but do not assume as in the preceding section that the paths are necessarily continuous. Assume, however, that $[W_t: t \geq 0]$ is separable with respect to D . Fix t_0 and δ . Choose sets $D_m = \{t_{m1}, \dots, t_{mm}\}$ of D -points such that $t_0 < t_{m1} < \dots < t_{mm} < t_0 + \delta$ and $D_m \uparrow D \cap (t_0, t_0 + \delta)$. By the argument leading to (37.9),

$$(38.6) \quad P \left[\sup_{\substack{t_0 \leq t \leq t_0 + \delta \\ t \in D}} |W_t - W_{t_0}| > \alpha \right] \leq \frac{K\delta^2}{\alpha^4}.$$

For sample points outside the N in the definition of separability, the supremum in (38.6) is unaltered, because of (38.5), if the restriction $t \in D$ is dropped. Since $P(N) = 0$,

$$P \left[\sup_{t_0 \leq t \leq t_0 + \delta} |W_t - W_{t_0}| > \alpha \right] \leq \frac{K\delta^2}{\alpha^4}.$$

Define M_n by (37.8) but with r ranging over all the reals (not just over the dyadic rationals) in $[k2^{-n}, (k+2)2^{-n}]$. Then $P[M_n > n^{-1}] \leq 4Kn^5/2^n$ follows just as before. But for ω outside $B = [M_n > n^{-1} \text{ i.o.}]$, $W(\cdot, \omega)$ is continuous. Since $P(B) = 0$, $W(\cdot, \omega)$ is continuous for ω outside an \mathcal{F} -set of probability 0. If (Ω, \mathcal{F}, P) is complete, then the set of ω for which $W(\cdot, \omega)$ is continuous is an \mathcal{F} -set of probability 1. Thus paths are continuous with probability 1 for any separable process having the finite-dimensional distributions of Brownian motion—provided that the underlying space is complete, which can of course always be arranged. ■

As it will be shown below that there exists a separable process with any consistently prescribed set of finite-dimensional distributions, Example 38.4 provides another approach to the construction of continuous Brownian motion. The value of the method lies in its generality. It must not, however, be imagined that separability automatically ensures smooth sample paths:

Example 38.5. Suppose that the random variables X_t , $t \geq 0$, are independent, each having the standard normal distribution. Let D be any countable set dense in $T = [0, \infty)$. Suppose that I and J are open intervals with rational endpoints. Since the random variables X_t with $t \in D \cap I$ are independent, and since the value common to the $P[X_t \in J]$ is positive, the second Borel–Cantelli lemma implies that with probability 1, $X_t \in J$ for some t in $D \cap I$. Since there are only countably many pairs I and J with rational endpoints, there is an \mathcal{F} -set N such that $P(N) = 0$ and such that for ω outside N the set $\{X(t, \omega): t \in D \cap I\}$ is everywhere dense on the line for every open interval I in T . This implies that $[X_t: t \geq 0]$ is separable with respect to D . But also of course it implies that the paths are highly irregular.

This irregularity is not a shortcoming of the concept of separability—it is a necessary consequence of the properties of the finite-dimensional distributions specified in this example. ■

Example 38.6. The process $[Y_t: t \geq 0]$ in Example 38.1 is not separable: The path $Y(\cdot, \omega)$ is not separable D unless D contains the point $V(\omega)$. The set of ω for which $Y(\cdot, \omega)$ is separable D is thus contained in $[\omega: V(\omega) \in D]$, a set of probability 0, since D is countable and V has a continuous distribution. ■

Existence Theorems

It will be proved in stages that for every consistent system of finite-dimensional distributions there exists a separable process having those distributions. Define x to be separable D at the point t if there exist points t_n in D such that $t_n \rightarrow t$ and $x(t_n) \rightarrow x(t)$. Note that this is no restriction on x if t lies in D , and note that separability is the same thing as separability at every t .

Lemma 1. Let $[X_t: t \geq 0]$ be a stochastic process on (Ω, \mathcal{F}, P) . There exists a countable, dense set D in $[0, \infty)$, and there exists for each t an \mathcal{F} -set $N(t)$, such that $P(N(t)) = 0$ and such that for ω outside $N(t)$ the path function $X(\cdot, \omega)$ is separable D at t .

PROOF. Fix open intervals I and J , and consider the probability

$$p(U) = P\left(\bigcap_{s \in U} [X_s \notin J]\right)$$

for countable subsets U of $I \cap T$. As U increases, the intersection here decreases and so does $p(U)$. Choose U_n so that $p(U_n) \rightarrow \inf_U p(U)$. If $U(I, J) = \bigcup_n U_n$, then $U(I, J)$ is a countable subset of $I \cap T$ making $p(U)$ minimal:

$$(38.7) \quad P\left(\bigcap_{s \in U(I, J)} [X_s \notin J]\right) \leq P\left(\bigcap_{s \in U} [X_s \notin J]\right)$$

for every countable subset U of $I \cap T$. If $t \in I \cap T$, then

$$(38.8) \quad P\left([X_t \in J] \cap \bigcap_{s \in U(I, J)} [X_s \notin J]\right) = 0,$$

because otherwise (38.7) would fail for $U = U(I, J) \cup \{t\}$.

Let $D = \bigcup U(I, J)$, where the union extends over all open intervals I and J with rational endpoints. Then D is a countable, dense subset of T . For each t let

$$(38.9) \quad N(t) = \bigcup \left([X_t \in J] \cap \bigcap_{s \in U(I, J)} [X_s \notin J] \right),$$

where the union extends over all open intervals J that have rational endpoints and over all open intervals I that have rational endpoints and contain t . Then $N(t)$ is by (38.8) an \mathcal{F} -set such that $P(N(t)) = 0$.

Fix t and $\omega \notin N(t)$. The problem is to show that $X(\cdot, \omega)$ is separable with respect to D at t . Given n , choose open intervals I and J that have rational endpoints and lengths less than n^{-1} and satisfy $t \in I$ and $X(t, \omega) \in J$. Since ω lies outside (38.9), there must be an s_n in $U(I, J)$ such that $X(s_n, \omega) \in J$. But then $s_n \in D$, $|s_n - t| < n^{-1}$, and $|X(s_n, \omega) - X(t, \omega)| < n^{-1}$. Thus $s_n \rightarrow t$ and $X(s_n, \omega) \rightarrow X(t, \omega)$ for a sequence s_1, s_2, \dots in D . ■

For any countable D , the set of ω for which $X(\cdot, \omega)$ is separable with respect to D at t is

$$(38.10) \quad \bigcap_{n=1}^{\infty} \bigcup_{\substack{|s-t| < n^{-1} \\ s \in D}} [\omega: |X(t, \omega) - X(s, \omega)| < n^{-1}].$$

This set lies in \mathcal{F} for each t , and the point of the lemma is that it is possible to choose D in such a way that each of these sets has probability 1.

Lemma 2. *Let $[X_t: t \geq 0]$ be a stochastic process on (Ω, \mathcal{F}, P) . Suppose that for all t and ω*

$$(38.11) \quad a < X(t, \omega) < b.$$

Then there exists on (Ω, \mathcal{F}, P) a process $[X'_t: t \geq 0]$ having these three properties:

- (i) $P[X'_t = X_t] = 1$ for each t .
- (ii) For some countable, dense subset D of $[0, \infty)$, $X'(\cdot, \omega)$ is separable D for every ω in Ω .
- (iii) For all t and ω ,

$$(38.12) \quad a \leq X'(t, \omega) \leq b.$$

PROOF. Choose a countable, dense set D and \mathcal{F} -sets $N(t)$ of probability 0 as in Lemma 1. If $t \in D$ or if $\omega \notin N(t)$, define $X'(t, \omega) = X(t, \omega)$. If $t \notin D$,

fix some sequence $\{s_n^{(t)}\}$ in D for which $\lim_n s_n^{(t)} = t$, and define $X'(t, \omega) = \limsup_n X(s_n^{(t)}, \omega)$ for $\omega \in N(t)$. To sum up,

$$(38.13) \quad X'(t, \omega) = \begin{cases} X(t, \omega) & \text{if } t \in D \text{ or } \omega \notin N(t), \\ \limsup_n X(s_n^{(t)}, \omega) & \text{if } t \notin D \text{ and } \omega \in N(t). \end{cases}$$

Since $N(t) \in \mathcal{F}$, X'_t is measurable \mathcal{F} for each t . Since $P(N(t)) = 0$, $P[X_t = X'_t] = 1$ for each t .

Fix t and ω . If $t \in D$, then certainly $X'(\cdot, \omega)$ is separable D at t , and so assume $t \notin D$. If $\omega \notin N(t)$, then by the construction of $N(t)$, $X(\cdot, \omega)$ is separable with respect to D at t , so that there exist points s_n in D such that $s_n \rightarrow t$ and $X(s_n, \omega) \rightarrow X(t, \omega)$. But $X(s_n, \omega) = X'(s_n, \omega)$ because $s_n \in D$, and $X(t, \omega) = X'(t, \omega)$ because $\omega \notin N(t)$. Hence $X'(s_n, \omega) \rightarrow X'(t, \omega)$, and so $X'(\cdot, \omega)$ is separable with respect to D at t . Finally, suppose that $t \notin D$ and $\omega \in N(t)$. Then $X'(t, \omega) = \lim_k X(s_{n_k}^{(t)}, \omega)$ for some sequence $\{n_k\}$ of integers. As $k \rightarrow \infty$, $s_{n_k}^{(t)} \rightarrow t$ and $X'(s_{n_k}^{(t)}, \omega) = X(s_{n_k}^{(t)}, \omega) \rightarrow X'(t, \omega)$, so that again $X'(\cdot, \omega)$ is separable with respect to D at t . Clearly, (38.11) implies (38.12). ■

Example 38.7. One must allow for the possibility of equality in (38.12). Suppose that $V(\omega) > 0$ for all ω and that V has a continuous distribution. Define

$$f(t) = \begin{cases} e^{-|t|} & \text{if } t \neq 0, \\ 0 & \text{if } t = 0, \end{cases}$$

and put $X(t, \omega) = f(t - V(\omega))$. If $[X'_t: t \geq 0]$ is any separable process with the same finite-dimensional distributions as $[X_t: t \geq 0]$, then $X'(\cdot, \omega)$ must with probability 1 assume the value 1 somewhere. In this case (38.11) holds for $a < 0$ and $b = 1$, and equality in (38.12) cannot be avoided. ■

If

$$(38.14) \quad \sup_{t, \omega} |X(t, \omega)| < \infty,$$

then (38.11) holds for some a and b . To treat the case in which (38.14) fails, it is necessary to allow for the possibility of infinite values. If $x(t)$ is ∞ or $-\infty$, replace the third condition in (38.2) by $x(t_n) \rightarrow \infty$ or $x(t_n) \rightarrow -\infty$. This extends the definition of separability to functions x that may assume infinite values and to processes $[X_t: t \geq 0]$ for which $X(t, \omega) = \pm\infty$ is a possibility.

Theorem 38.1. *If $[X_t: t \geq 0]$ is a finite-valued process on (Ω, \mathcal{F}, P) , there exists on the same space a separable process $[X'_t: t \geq 0]$ such that $P[X'_t = X_t] = 1$ for each t .*

It is assumed for convenience here that $X(t, \omega)$ is finite for all t and ω , although this is not really necessary. But in some cases infinite values for certain $X'(t, \omega)$ cannot be avoided—see Example 38.8.

PROOF. If (38.14) holds, the result is an immediate consequence of Lemma 2. The definition of separability allows an exceptional set N of probability 0; in the construction of Lemma 2 this set is actually empty, but it is clear from the definition this could be arranged anyway.

The case in which (38.14) may fail could be treated by tracing through the preceding proofs, making slight changes to allow for infinite values. A simple argument makes this unnecessary. Let g be a continuous, strictly increasing mapping of R^1 onto $(0, 1)$. Let $Y(t, \omega) = g(X(t, \omega))$. Lemma 2 applies to $[Y_t: t \geq 0]$; there exists a separable process $[Y'_t: t \geq 0]$ such that $P[Y'_t = Y_t] = 1$. Since $0 < Y(t, \omega) < 1$, Lemma 2 ensures $0 \leq Y'(t, \omega) \leq 1$. Define

$$X'(t, \omega) = \begin{cases} -\infty & \text{if } Y'(t, \omega) = 0, \\ g^{-1}(Y'(t, \omega)) & \text{if } 0 < Y'(t, \omega) < 1, \\ +\infty & \text{if } Y'(t, \omega) = 1. \end{cases}$$

Then $[X'_t: t \geq 0]$ satisfies the requirements. Note that $P[X'_t = \pm\infty] = 0$ for each t . ■

Example 38.8. Suppose that $V(\omega) > 0$ for all ω and V has a continuous distribution. Define

$$h(t) = \begin{cases} |t|^{-1} & \text{if } t \neq 0, \\ 0 & \text{if } t = 0, \end{cases}$$

and put $X(t, \omega) = h(t - V(\omega))$. This is analogous to Example 38.7. If $[X'_t: t \geq 0]$ is separable and has the finite-dimensional distributions of $[X_t: t \geq 0]$, then $X'(\cdot, \omega)$ must with probability 1 assume the value ∞ for some t . ■

Combining Theorem 38.1 with Kolmogorov's existence theorem shows that for any consistent system of finite-dimensional distributions μ_{t_1, \dots, t_k} there exists a separable process with the μ_{t_1, \dots, t_k} as finite-dimensional distributions. As shown in Example 38.4, this leads to another construction of Brownian motion with continuous paths.

Consequences of Separability

The next theorem implies in effect that, if the finite-dimensional distributions of a process are such that it “should” have continuous paths, then it will in fact have continuous paths if it is separable. Example 38.4 illustrates this. The same thing holds for properties other than continuity.

Let \bar{R}^T be the set of functions on $T = [0, \infty)$ with values that are ordinary reals or else ∞ or $-\infty$. Thus \bar{R}^T is an enlargement of the R^T of Section 36, an enlargement necessary because separability sometimes forces infinite values. Define the function Z_t on \bar{R}^T by $Z_t(x) = Z(t, x) = x(t)$. This is just an extension of the coordinate function (36.8). Let $\bar{\mathcal{R}}^T$ be the σ -field in \bar{R}^T generated by the Z_t , $t \geq 0$.

Suppose that A is a subset of \bar{R}^T , not necessarily in $\bar{\mathcal{R}}^T$. For $D \subset T = [0, \infty)$, let A_D consist of those elements x of \bar{R}^T that agree on D with some element y of A :

$$(38.15) \quad A_D = \bigcup_{y \in A} \bigcap_{t \in D} [x \in \bar{R}^T: x(t) = y(t)].$$

Of course, $A \subset A_D$. Let S_D denote the set of x in \bar{R}^T that are separable with respect to D .

In the following theorem, $[X_t: t \geq 0]$ and $[X'_t: t \geq 0]$ are processes on spaces (Ω, \mathcal{F}, P) and $(\Omega', \mathcal{F}', P')$, which may be distinct; the path functions are $X(\cdot, \omega)$ and $X'(\cdot, \omega')$.

Theorem 38.2. *Suppose of A that for each countable, dense subset D of $T = [0, \infty)$, the set (38.15) satisfies*

$$(38.16) \quad A_D \in \bar{\mathcal{R}}^T, \quad A_D \cap S_D \subset A.$$

If $[X_t: t \geq 0]$ and $[X'_t: t \geq 0]$ have the same finite-dimensional distributions, if $[\omega: X(\cdot, \omega) \in A]$ lies in \mathcal{F} and has P -measure 1, and if $[X'_t: t \geq 0]$ is separable, then $[\omega': X'(\cdot, \omega') \in A]$ contains an \mathcal{F}' -set of P' -measure 1.

If $(\Omega', \mathcal{F}', P')$ is complete, then of course $[\omega': X'(\cdot, \omega') \in A]$ is itself an \mathcal{F}' -set of P' -measure 1.

PROOF. Suppose that $[X'_t: t \geq 0]$ is separable with respect to D . The difference $[\omega': X'(\cdot, \omega') \in A_D] - [\omega': X'(\cdot, \omega') \in A]$ is by (38.16) a subset of $[\omega': X'(\cdot, \omega') \in \bar{R}^T - S_D]$, which is contained in an \mathcal{F}' -set of N' of P' -measure 0. Since the two processes have the same finite-dimensional distributions and hence induce the same distribution on $(\bar{R}^T, \bar{\mathcal{R}}^T)$, and since A_D lies in $\bar{\mathcal{R}}^T$, it follows that $P'[\omega': X'(\cdot, \omega') \in A_D] = P[\omega: X(\cdot, \omega) \in A_D] \geq P[\omega: X(\cdot, \omega) \in A] = 1$. Thus the subset $[\omega': X'(\cdot, \omega') \in A_D] - N'$ of $[\omega': X'(\cdot, \omega') \in A]$ lies in \mathcal{F}' and has P' -measure 1. ■

Example 38.9. Consider the set C of finite-valued, continuous functions on T . If $x \in S_D$ and $y \in C$, and if x and y agree on a dense D , then x and y agree everywhere: $x = y$. Therefore, $C_D \cap S_D \subset C$. Further,

$$C_D = \bigcap_{\epsilon, t} \bigcup_{\delta} \bigcap_s [x \in \bar{R}^T: |x(s)| < \infty, |x(t)| < \infty, |x(s) - x(t)| < \epsilon],$$

where ϵ and δ range over the positive rationals, t ranges over D , and the inner intersection extends over the s in D satisfying $|s - t| < \delta$. Hence $C_D \in \overline{\mathcal{R}}^T$. Thus C satisfies the condition (38.16).

Theorem 38.2 now implies that if a process has continuous paths with probability 1, then any separable process having the same finite-dimensional distributions has continuous paths outside a set of probability 0. In particular, a Brownian motion with continuous paths was constructed in the preceding section, and so any separable process with the finite-dimensional distributions of Brownian motion has continuous paths outside a set of probability 0. The argument in Example 38.4 now becomes supererogatory. ■

Example 38.10. There is a somewhat similar argument for the step functions of the Poisson process. Let Z^+ be the set of nonnegative integers; let E consist of the nondecreasing functions x in \overline{R}^T such that $x(t) \in Z^+$ for all t and such that for every $n \in Z^+$ there exists a nonempty interval I such that $x(t) = n$ for $t \in I$. Then

$$E_D = \bigcap_{t \in D} [x: x(t) \in Z^+] \cap \bigcap_{s, t \in D, s < t} [x: x(s) \leq x(t)] \\ \cap \bigcap_{n=0}^{\infty} \bigcup_I \bigcap_{t \in D \cap I} [x: x(t) = n],$$

where I ranges over the open intervals with rational endpoints. Thus $E_D \in \overline{\mathcal{R}}^T$. Clearly, $E_D \cap S_D \subset E$, and so Theorem 38.2 applies.

In Section 23 was constructed a Poisson process with paths in E , and therefore any separable process with the same finite-dimensional distributions will have paths in E except for a set of probability 0. ■

Example 38.11. For E as in Example 38.10, let E_0 consist of the elements of E that are right-continuous; a function in E need not lie in E_0 , although at each t it must be continuous from one side or the other. The Poisson process as defined in Section 23 by $N_t = \max\{n: S_n \leq t\}$ (see (23.5)) has paths in E_0 . But if $N'_t = \max\{n: S_n < t\}$, then $[N'_t: t \geq 0]$ is separable and has the same finite-dimensional distributions, but its paths are not in E_0 . Thus E_0 does not satisfy the hypotheses of Theorem 38.2. Separability does not help distinguish between continuity from the right and continuity from the left. ■

Example 38.12. The class of sets A satisfying (38.16) is closed under the formation of countable unions and intersections but is not closed under complementation. Define X_t and Y_t as in Example 38.1, and let C be the set of continuous paths. Then $[Y_t: t \geq 0]$ and $[X_t: t \geq 0]$ have the same finite-dimensional distributions, and the latter is separable; $Y(\cdot, \omega)$ is in $\overline{R}^T - C$ for each ω , and $X(\cdot, \omega)$ is in $\overline{R}^T - C$ for no ω . ■

Example 38.13. As a final example, consider the set J of functions with discontinuities of at most the first kind: x is in J if it is finite-valued, if $x(t+) = \lim_{s \downarrow t} x(s)$ exists (finite) for $t \geq 0$ and $x(t-) = \lim_{s \uparrow t} x(s)$ exists (finite) for $t > 0$, and if $x(t)$ lies between $x(t+)$ and $x(t-)$ for $t > 0$. Continuous and right-continuous functions are special cases.

Let V denote the general system

$$(38.17) \quad V: k; r_1, \dots, r_k; s_1, \dots, s_k; \alpha_1, \dots, \alpha_k,$$

where k is an integer, where the r_i , s_i , and α_i are rational, and where

$$0 = r_1 < s_1 < r_2 < s_2 < \dots < r_k < s_k.$$

Define

$$J(D, V, \epsilon) = \bigcap_{i=1}^k [x: \alpha_i \leq x(t) \leq \alpha_i + \epsilon, t \in (r_i, s_i) \cap D] \\ \cap \bigcap_{i=2}^k [x: \min\{\alpha_{i-1}, \alpha_i\} \leq x(t) \leq \max\{\alpha_{i-1}, \alpha_i\} + \epsilon, t \in (s_{i-1}, r_i) \cap D].$$

Let $\mathcal{U}_{m,k,\delta}$ be the class of systems (38.17) that have a fixed value for k and satisfy $r_i - s_{i-1} < \delta$, $i = 2, \dots, k$, and $s_k > m$. It will be shown that

$$(38.18) \quad J_D = \bigcap_{m=1}^{\infty} \bigcap_{\epsilon} \bigcup_{k=1}^{\infty} \bigcap_{\delta} \bigcup_{V \in \mathcal{U}_{m,k,\delta}} J(D, V, \epsilon),$$

where ϵ and δ range over the positive rationals. From this it will follow that $J_D \in \overline{\mathcal{R}}^T$. It will also be shown that $J_D \cap S_D \subset J$, so that J satisfies the hypothesis of Theorem 38.2.

Suppose that $y \in J$. For fixed ϵ , let H be the set of nonnegative h for which there exist finitely many points t_i such that $0 = t_0 \leq t_1 \leq \dots \leq t_r = h$ and $|y(t) - y(t')| < \epsilon$ for t and t' in the same interval (t_{i-1}, t_i) . If $h_n \in H$ and $h_n \uparrow h$, then from the existence of $y(h-)$ follows $h \in H$. Hence H is closed. If $h \in H$, from the existence of $y(h+)$ it follows that H contains points to the right of h . Therefore, $H = [0, \infty)$. From this it follows that the right side of (38.18) contains J_D .

Suppose that x is a member of the right side of (38.18). It is not hard to deduce that for each t the limits

$$(38.19) \quad \lim_{s \downarrow t, s \in D} x(s), \quad \lim_{s \uparrow t, s \in D} x(s)$$

exist and that $x(t)$ lies between them if $t \in D$. For $t \in D$ take $y(t) = x(t)$, and for $t \notin D$ take $y(t)$ to be the first limit in (38.19). Then $y \in J$ and hence $x \in J_D$. This argument also shows that $J_D \cap S_D \subset J$. ■

Appendix

Gathered here for easy reference are certain definitions and results from set theory and real analysis required in the text. Although there are many newer books, HAUSDORFF (the early sections) on set theory and HARDY on analysis are still excellent for the general background assumed here.

Set Theory

A1. The *empty set* is denoted by \emptyset . *Sets* are variable subsets of some *space* that is fixed in any one definition, argument, or discussion; this space is denoted either generically by Ω or by some special symbol (such as R^k for Euclidean k -space). A *singleton* is a set consisting of just one point or element. That A is a *subset* of B is expressed by $A \subset B$. In accordance with standard usage, $A \subset B$ does not preclude $A = B$; A is a *proper subset* of B if $A \subset B$ and $A \neq B$.

The *complement* of A is always relative to the overall space Ω ; it consists of the points of Ω not contained in A and is denoted by A^c . The *difference* between A and B , denoted by $A - B$, is $A \cap B^c$; here B need not be contained in A , and if it is, then $A - B$ is a *proper difference*. The *symmetric difference* $A \triangle B = (A \cap B^c) \cup (A^c \cap B)$ consists of the points that lie in one of the sets A and B but not in both.

Classes of sets are denoted by script letters. The *power set* of Ω is the class of all subsets of Ω ; it is denoted 2^Ω .

A2. The set of ω that lie in A and satisfy a given property $p(\omega)$ is denoted $[\omega \in A: p(\omega)]$; if $A = \Omega$, this is usually shortened to $[\omega: p(\omega)]$.

A3. In this book, to say that a collection $[A_\theta: \theta \in \Theta]$ is *disjoint* always means that it is *pairwise disjoint*: $A_\theta \cap A_{\theta'} = \emptyset$ if θ and θ' are distinct elements of the index set Θ . To say that A *meets* B , or that B *meets* A , is to say that they are not disjoint: $A \cap B \neq \emptyset$. The collection $[A_\theta: \theta \in \Theta]$ *covers* B if $B \subset \bigcup_\theta A_\theta$. The collection is a *decomposition* or *partition* of B if it is disjoint and $B = \bigcup_\theta A_\theta$.

A4. By $A_n \uparrow A$ is meant $A_1 \subset A_2 \subset \cdots$ and $A = \bigcup_n A_n$; by $A_n \downarrow A$ is meant $A_1 \supset A_2 \supset \cdots$ and $A = \bigcap_n A_n$.

A5. The *indicator*, or *indicator function*, of a set A is the function on Ω that assumes the value 1 on A and 0 on A^c ; it is denoted I_A . The alternative term “characteristic function” is reserved for the Fourier transform (see Section 26).

A6. *De Morgan's laws* are $(\bigcup_{\theta} A_{\theta})^c = \bigcap_{\theta} A_{\theta}^c$ and $(\bigcap_{\theta} A_{\theta})^c = \bigcup_{\theta} A_{\theta}^c$. These and the other facts of basic set theory are assumed known: a countable union of countable sets is countable, and so on.

A7. If $T: \Omega \rightarrow \Omega'$ is a mapping of Ω into Ω' and A' is a set in Ω' , the *inverse image* of A' is $T^{-1}A' = [\omega \in \Omega: T\omega \in A']$. It is easily checked that each of these statements is equivalent to the next: $\omega \in \Omega - T^{-1}A'$, $\omega \notin T^{-1}A'$, $T\omega \notin A'$, $T\omega \in \Omega' - A'$, $\omega \in T^{-1}(\Omega' - A')$. Therefore, $\Omega - T^{-1}A' = T^{-1}(\Omega' - A')$. Simple considerations of this kind show that $\bigcup_{\theta} T^{-1}A'_{\theta} = T^{-1}(\bigcup_{\theta} A'_{\theta})$ and $\bigcap_{\theta} T^{-1}A'_{\theta} = T^{-1}(\bigcap_{\theta} A'_{\theta})$, and that $A' \cap B' = \emptyset$ implies $T^{-1}A' \cap T^{-1}B' = \emptyset$ (the reverse implication is false unless $T\Omega = \Omega'$).

If f maps Ω into another space, $f(\omega)$ is the value of the function f at an unspecified value of the argument ω . The function f itself (the rule defining the mapping) is sometimes denoted $f(\cdot)$. This is especially convenient for a function $f(\omega, t)$ of two arguments: For each fixed t , $f(\cdot, t)$ denotes the function on Ω with value $f(\omega, t)$ at ω .

A8. *The axiom of choice.* Suppose that $[A_{\theta}: \theta \in \Theta]$ is a decomposition of Ω into nonempty sets. The axiom of choice says that there exists a set (at least one set) C that contains exactly one point from each A_{θ} : $C \cap A_{\theta}$ is a singleton for each θ in Θ . The existence of such sets C is assumed in "everyday" mathematics, and the axiom of choice may even seem to be simply *true*. A careful treatment of set theory, however, is based on an explicit list of such axioms and a study of the relationships between them; see HALMOS₂ or DUDLEY.

A few of the problems require *Zorn's lemma*, which is equivalent to the axiom of choice; see DUDLEY or KAPLANSKY.

The Real Line

A9. The real line is denoted by R^1 ; $x \vee y = \max\{x, y\}$ and $x \wedge y = \min\{x, y\}$. For real x , $[x]$ is the integer part of x , and $\text{sgn } x$ is $+1$, 0 , or -1 as x is positive, 0 , or negative. It is convenient to be explicit about open, closed, and half-open intervals:

$$(a, b) = [x: a < x < b],$$

$$[a, b] = [x: a \leq x \leq b],$$

$$(a, b] = [x: a < x \leq b],$$

$$[a, b) = [x: a \leq x < b].$$

A10. Of course $x_n \rightarrow x$ means $\lim_n x_n = x$; $x_n \uparrow x$ means $x_1 \leq x_2 \leq \cdots$ and $x_n \rightarrow x$; $x_n \downarrow x$ means $x_1 \geq x_2 \geq \cdots$ and $x_n \rightarrow x$.

A sequence $\{x_n\}$ is *bounded* if and only if every subsequence $\{x_{n_k}\}$ contains a further subsequence $\{x_{n_{k(j)}}\}$ that converges to some x : $\lim_j x_{n_{k(j)}} = x$. If $\{x_n\}$ is not bounded, then for each k there is an n_k for which $|x_{n_k}| > k$; no subsequence of $\{x_{n_k}\}$ can converge. The implication in the other direction is a simple consequence of the fact that every bounded sequence contains a convergent subsequence.

If $\{x_n\}$ is bounded, and if each subsequence that converges at all converges to x , then $\lim_n x_n = x$. If x_n does not converge to x , then $|x_{n_k} - x| > \epsilon$ for some positive ϵ and some increasing sequence $\{n_k\}$ of integers; some subsequence of $\{x_{n_k}\}$ converges, but the limit cannot be x .

A11. A set G is defined as *open* if for each x in G there is an open interval I such that $x \in I \subset G$. A set F is defined as *closed* if F^c is open. The *interior* of A , denoted

A° , consists of the x in A for which there exists an open interval I such that $x \in I \subset A$. The *closure* of A , denoted A^- , consists of the x for which there exists a sequence $\{x_n\}$ in A with $x_n \rightarrow x$. The *boundary* of A is $\partial A = A^- - A^\circ$. The basic facts of real analysis are assumed known: A is open if and only if $A = A^\circ$; A is closed if and only if $A = A^-$; A is closed if and only if it contains all limits of sequences in it; x lies in ∂A if and only if there is a sequence $\{x_n\}$ in A and a sequence $\{y_n\}$ in A^c such that $x_n \rightarrow x$ and $y_n \rightarrow x$; and so on.

A12. Every open set G on the line is a countable, disjoint union of open intervals. To see this, define points x and y of G to be equivalent if $x \leq y$ and $[x, y] \subset G$ or $y \leq x$ and $[y, x] \subset G$. This is an equivalence relation. Each equivalence class is an interval, and since G is open, each is in fact an open interval. Thus G is a disjoint union of open (nonempty) intervals, and there can be only countably many of them, since each contains a rational.

A13. The simplest form of the *Heine–Borel* theorem says that if $[a, b] \subset \bigcup_{k=1}^{\infty} (a_k, b_k)$, then $[a, b] \subset \bigcup_{k=1}^n (a_k, b_k)$ for some n . A set A is defined to be *compact* if each cover of it by open sets has a finite subcover—that is, if $\{G_\theta: \theta \in \Theta\}$ covers A and each G_θ is open, then some finite subcollection $\{G_{\theta_1}, \dots, G_{\theta_n}\}$ covers A . Equivalent to the Heine–Borel theorem is the assertion that a bounded, closed set is compact. Also equivalent is the assertion that every bounded sequence of real numbers has a convergent subsequence.

A14. *The diagonal method.* From this last fact follows one of the basic principles of analysis.

Theorem. Suppose that each row of the array

$$(1) \quad \begin{array}{cccc} x_{1,1} & x_{1,2} & x_{1,3} & \cdots \\ x_{2,1} & x_{2,2} & x_{2,3} & \cdots \\ \vdots & \vdots & \vdots & \end{array}$$

is a bounded sequence of real numbers. Then there exists an increasing sequence n_1, n_2, \dots of integers such that the limit $\lim_k x_{r, n_k}$ exists for $r = 1, 2, \dots$.

PROOF. From the first row, select a convergent subsequence

$$(2) \quad x_{1, n_{1,1}}, x_{1, n_{1,2}}, x_{1, n_{1,3}}, \dots;$$

here $\{n_{1,k}\}$ is an increasing sequence of integers and $\lim_k x_{1, n_{1,k}}$ exists. Look next at the second row of (1) along the sequence $n_{1,1}, n_{1,2}, \dots$:

$$(3) \quad x_{2, n_{1,1}}, x_{2, n_{1,2}}, x_{2, n_{1,3}}, \dots$$

As a subsequence of the second row of (1), (3) is bounded. Select from it a convergent subsequence

$$x_{2, n_{2,1}}, x_{2, n_{2,2}}, x_{2, n_{2,3}}, \dots;$$

here $\{n_{2,k}\}$ is an increasing sequence of integers, a subsequence of $\{n_{1,k}\}$, and $\lim_k x_{2, n_{2,k}}$ exists.

Continue inductively in the same way. This gives an array

$$(4) \quad \begin{array}{cccc} n_{1,1} & n_{1,2} & n_{1,3} & \cdots \\ n_{2,1} & n_{2,2} & n_{2,3} & \cdots \\ \vdots & \vdots & \vdots & \end{array}$$

with three properties: (i) Each row of (4) is an increasing sequence of integers. (ii) The r th row is a subsequence of the $(r-1)$ st. (iii) For each r , $\lim_k x_{r,n_{r,k}}$ exists. Thus

$$(5) \quad x_{r,n_{r,1}}, x_{r,n_{r,2}}, x_{r,n_{r,3}}, \dots$$

is a convergent subsequence of the r th row of (1).

Put $n_k = n_{k,k}$. Since each row of (4) is increasing and is contained in the preceding row, n_1, n_2, n_3, \dots is an increasing sequence of integers. Furthermore, $n_r, n_{r+1}, n_{r+2}, \dots$ is a subsequence of the r th row of (4). Thus $x_{r,n_r}, x_{r,n_{r+1}}, x_{r,n_{r+2}}, \dots$ is a subsequence of (5) and is therefore convergent. Thus $\lim_k x_{r,n_k}$ does exist. ■

Since $\{n_k\}$ is the diagonal of the array (4), application of this theorem is called the *diagonal method*.

A15. The set A is by definition *dense* in the set B if for each x in B and each open interval J containing x , J meets A . This is the same thing as requiring $B \subset A^-$. The set E is by definition *nowhere dense* if each open interval I contains some open interval J that does not meet E . This makes sense: If I contains an open interval J that does not meet E , then E is not dense in I ; the definition requires that E be dense in *no* interval I .

A set A is defined to be *perfect* if it is closed and for each x in A and positive ϵ , there is a y in A such that $0 < |x - y| < \epsilon$. An equivalent requirement is that A be closed and for each x in A there exist a sequence $\{x_n\}$ in A such that $x_n \neq x$ and $x_n \rightarrow x$. The Cantor set is uncountable, nowhere dense, and perfect.

A set that is nowhere dense is in a sense small. If A is a countable union of sets each of which is nowhere dense, then A is said to be of the *first category*. This is a weaker notion of smallness. A set that is not of the first category is said to be of the *second category*.

Euclidean k -Space

A16. Euclidean space of dimension k is denoted R^k . Points (a_1, \dots, a_k) and (b_1, \dots, b_k) determine open, closed, and half-open rectangles in R^k :

$$\begin{aligned} [x: a_i < x_i < b_i, i = 1, \dots, k], \\ [x: a_i \leq x_i \leq b_i, i = 1, \dots, k], \\ [x: a_i < x_i \leq b_i, i = 1, \dots, k]. \end{aligned}$$

A rectangle (without a qualifying adjective) is in this book a set of this last form.

The Euclidean distance $(\sum_{i=1}^k (x_i - y_i)^2)^{1/2}$ between $x = (x_1, \dots, x_k)$ and $y = (y_1, \dots, y_k)$ is denoted by $|x - y|$.

A17. All the concepts in A11 carry over to R^k : simply take the I there to be an open rectangle in R^k . The definition of compact set also carries over word for word, and the Heine–Borel theorem in R^k says that a closed, bounded set is compact.

Analysis

A18. The standard Landau notation is used. Suppose that $\{x_n\}$ and $\{y_n\}$ are real sequences and $y_n > 0$. Then $x_n = O(y_n)$ means x_n/y_n is bounded; $x_n = o(y_n)$ means $x_n/y_n \rightarrow 0$; $x_n \sim y_n$ means $x_n/y_n \rightarrow 1$; $x_n \asymp y_n$ means x_n/y_n and y_n/x_n are both bounded. To write $x_n = z_n + O(y_n)$, for example, means that $x_n = z_n + u_n$ for some $\{u_n\}$ satisfying $u_n = O(y_n)$ —that is, that $x_n - z_n = O(y_n)$.

A19. *A difference equation.* Suppose that a and b are integers and $a < b$. Suppose that x_n is defined for $a \leq n \leq b$ and satisfies

$$(6) \quad x_n = px_{n+1} + qx_{n-1} \quad \text{for } a < n < b,$$

where p and q are positive and $p + q = 1$. The general solution of this difference equation has the form

$$(7) \quad x_n = \begin{cases} A + B(q/p)^n & \text{for } a \leq n \leq b \quad \text{if } p \neq q, \\ A + Bn & \text{for } a \leq n \leq b \quad \text{if } p = q. \end{cases}$$

That (7) always solves (6) is easily checked. Suppose the values x_{n_1} and x_{n_2} are given, where $a \leq n_1 < n_2 \leq b$. If $p \neq q$, the system

$$A + B(q/p)^{n_1} = x_{n_1}, \quad A + B(q/p)^{n_2} = x_{n_2}$$

can always be solved for A and B . If $p = q$, the system

$$A + Bn_1 = x_{n_1}, \quad A + Bn_2 = x_{n_2}$$

can always be solved. Take $n_1 = a$ and $n_2 = a + 1$; the corresponding A and B satisfy (7) for $n = a$ and for $n = a + 1$, and it follows by induction that (7) holds for all n . Thus any solution of (6) can indeed be put in the form (7). Furthermore, the equation (6) and any pair of values x_{n_1} and x_{n_2} ($n_1 \neq n_2$) suffice to determine all the x_n .

If x_n is defined for $a \leq n < \infty$ and satisfies (6) for $a < n < \infty$, then there are constants A and B such that (7) holds for $a \leq n < \infty$.

A20. *Cauchy's equation.*

Theorem. Let f be a real function on $(0, \infty)$, and suppose that f satisfies Cauchy's equation: $f(x + y) = f(x) + f(y)$ for $x, y > 0$. If there is some interval on which f is bounded above, then $f(x) = xf(1)$ for $x > 0$.

PROOF. The problem is to prove that $g(x) = f(x) - xf(1)$ vanishes identically. Clearly, $g(1) = 0$, and g satisfies Cauchy's equation and on some interval is bounded above. By induction, $g(nx) = ng(x)$; hence $ng(m/n) = g(m) = mg(1) = 0$, so that $g(r) = 0$ for positive rational r . Suppose that $g(x_0) \neq 0$ for some x_0 . If $g(x_0) < 0$, then $g(r_0 - x_0) = -g(x_0) > 0$ for rational $r_0 > x_0$. It is thus no restriction to assume that $g(x_0) > 0$. Let I be an open interval in which g is bounded above. Given a number M , choose n so that $ng(x_0) > M$, and then choose a rational r so that $nx_0 + r$ lies in I . If $r > 0$, then $g(r + nx_0) = g(r) + g(nx_0) = g(nx_0) = ng(x_0)$. If $r < 0$, then $ng(x_0) = g(nx_0) = g((-r) + (nx_0 + r)) = g(-r) + g(nx_0 + r) = g(nx_0 + r)$. In either

case, $g(nx_0 + r) = ng(x_0)$; of course this is trivial if $r = 0$. Since $g(nx_0 + r) = ng(x_0) > M$ and M was arbitrary, g is not bounded above in I , a contradiction. ■

Obviously, the same proof works if f is bounded below in some interval.

Corollary. *Let U be a real function on $(0, \infty)$ and suppose that $U(x + y) = U(x)U(y)$ for $x, y > 0$. Suppose further that there is some interval on which U is bounded above. Then either $U(x) = 0$ for $x > 0$, or else there is an A such that $U(x) = e^{Ax}$ for $x > 0$.*

PROOF. Since $U(x) = U^2(x/2)$, U is nonnegative. If $U(x) = 0$, then $U(x/2^n) = 0$ and so U vanishes at points arbitrarily near 0. If U vanishes at a point, it must by the functional equation vanish everywhere to the right of that point. Hence U is identically 0 or else everywhere positive.

In the latter case, the theorem applies to $f(x) = \log U(x)$, this function being bounded above in some interval, and so $f(x) = Ax$ for $A = \log U(1)$. ■

A21. A number-theoretic fact.

Theorem. *Suppose that M is a set of positive integers closed under addition and that M has greatest common divisor 1. Then M contains all integers exceeding some n_0 .*

PROOF. Let M_1 consist of all the integers m , $-m$, and $m - m'$ with m and m' in M . Then M_1 is closed under addition and subtraction (it is a subgroup of the group of integers). Let d be the smallest positive element of M_1 . If $n \in M_1$, write $n = qd + r$, where $0 \leq r < d$. Since $r = n - qd$ lies in M_1 , r must actually be 0. Thus M_1 consists of the multiples of d . Since d divides all the integers in M_1 and hence all the integers in M , and since M has greatest common divisor 1, $d = 1$. Thus M_1 contains all the integers.

Write $1 = m - m'$ with m and m' in M (if 1 itself is in M , the proof is easy), and take $n_0 = (m + m')^2$. Given $n > n_0$, write $n = q(m + m') + r$, where $0 \leq r < m + m'$. From $n > n_0 \geq (r + 1)(m + m')$ follows $q = (n - r)/(m + m') > r$. But $n = q(m + m') + r(m - m') = (q + r)m + (q - r)m'$, and since $q + r \geq q - r > 0$, n lies in M . ■

A22. One- and two-sided derivatives.

Theorem. *Suppose that f and g are continuous on $[0, \infty)$ and g is the right-hand derivative of f on $(0, \infty)$: $f^+(t) = g(t)$ for $t > 0$. Then $f^+(0) = g(0)$ as well, and g is the two-sided derivative of f on $(0, \infty)$.*

PROOF. It suffices to show that $F(t) = f(t) - f(0) - \int_0^t g(s) ds$ vanishes for $t \geq 0$. By assumption, F is continuous on $[0, \infty)$ and $F^+(t) = 0$ for $t > 0$. Suppose that $F(t_0) > F(t_1)$, where $0 < t_0 < t_1$. Then $G(t) = F(t) - (t - t_0)(F(t_1) - F(t_0))/(t_1 - t_0)$ is continuous on $[0, \infty)$, $G(t_0) = G(t_1)$, and $G^+(t) > 0$ on $(0, \infty)$. But then the maximum of G over $[t_0, t_1]$ must occur at some interior point; since $G^+ \leq 0$ at a local maximum, this is impossible. Similarly $F(t_0) < F(t_1)$ is impossible. Thus F is constant over $(0, \infty)$ and by continuity is constant over $[0, \infty)$. Since $F(0) = 0$, F vanishes on $[0, \infty)$. ■

A23. A differential equation. The equation $f'(t) = Af(t) + g(t)$ ($t \geq 0$; g continuous) has the particular solution $f_0(t) = e^{At} \int_0^t g(s) e^{-As} ds$; for an arbitrary solution f , $(f(t) - f_0(t))e^{-At}$ has derivative 0 and hence equals $f(0)$ identically. All solutions thus have the form $f(t) = e^{At}[f(0) + \int_0^t g(s) e^{-As} ds]$.

A24. *A trigonometric identity.* If $z \neq 1$ and $z \neq 0$, then

$$\sum_{k=-l}^l z^k = z^{-l} \sum_{k=0}^{2l} z^k = z^{-l} \frac{1 - z^{2l+1}}{1 - z} = \frac{z^{-l} - z^{l+1}}{1 - z},$$

and hence

$$\begin{aligned} \sum_{l=0}^{m-1} \sum_{k=-l}^l z^k &= \sum_{l=0}^{m-1} \frac{z^{-l} - z^{l+1}}{1 - z} = \frac{1}{1 - z} \left[\frac{1 - z^{-m}}{1 - z^{-1}} - z \frac{1 - z^m}{1 - z} \right] \\ &= \frac{1 - z^{-m} + 1 - z^m}{(1 - z)(1 - z^{-1})} = \frac{(z^{m/2} - z^{-m/2})^2}{(z^{1/2} - z^{-1/2})^2}. \end{aligned}$$

Take $z = e^{ix}$. If x is not an integral multiple of 2π , then

$$\sum_{l=0}^{m-1} \sum_{k=-l}^l e^{ikx} = \frac{(\sin \frac{1}{2}mx)^2}{(\sin \frac{1}{2}x)^2}.$$

If $x = 2\pi n$, the left-hand side here is m^2 , which is the limit of the right-hand side as $x \rightarrow 2\pi n$.

Infinite Series

A25. *Nonnegative series.* Suppose x_1, x_2, \dots are nonnegative. If E is a finite set of integers, then $E \subset \{1, 2, \dots, n\}$ for some n , so that by nonnegativity $\sum_{k \in E} x_k \leq \sum_{k=1}^n x_k$. The set of partial sums $\sum_{k=1}^n x_k$ thus has the same supremum as the larger set of sums $\sum_{k \in E} x_k$ (E finite). Therefore, the nonnegative series $\sum_{k=1}^\infty x_k$ converges if and only if the sums $\sum_{k \in E} x_k$ for finite E are bounded, in which case the sum is the supremum: $\sum_{k=1}^\infty x_k = \sup_E \sum_{k \in E} x_k$.

A26. *Dirichlet's theorem.* Since the supremum in A25 is invariant under permutations, so is $\sum_{k=1}^\infty x_k$: If the x_k are nonnegative and $y_k = x_{f(k)}$ for some one-to-one map f of the positive integers onto themselves, then $\sum_k x_k$ and $\sum_k y_k$ diverge or converge together and in the latter case have the same sum.

A27. *Double series.* Suppose that x_{ij} , $i, j = 1, 2, \dots$, are nonnegative. The i th row gives a series $\sum_j x_{ij}$, and if each of these converges, one can form the series $\sum_i \sum_j x_{ij}$. Let the terms x_{ij} be arranged in some order as a single infinite series $\sum_{ij} x_{ij}$; by Dirichlet's theorem, the sum is the same whatever order is used.

Suppose each $\sum_j x_{ij}$ converges and $\sum_i \sum_j x_{ij}$ converges. If E is a finite set of the pairs (i, j) , there is an n for which $\sum_{(i,j) \in E} x_{ij} \leq \sum_{i \leq n} \sum_{j \leq n} x_{ij} \leq \sum_{i \leq n} \sum_j x_{ij} \leq \sum_i \sum_j x_{ij}$; hence $\sum_{ij} x_{ij}$ converges and has sum at most $\sum_i \sum_j x_{ij}$. On the other hand, if $\sum_{ij} x_{ij}$ converges, then $\sum_{i \leq m} \sum_{j \leq n} x_{ij} \leq \sum_{ij} x_{ij}$; letting $n \rightarrow \infty$ and then $m \rightarrow \infty$ shows that each $\sum_j x_{ij}$ converges and that $\sum_i \sum_j x_{ij} \leq \sum_{ij} x_{ij}$. Therefore, in the nonnegative case, $\sum_{ij} x_{ij}$ converges if and only if the $\sum_j x_{ij}$ all converge and $\sum_i \sum_j x_{ij}$ converges, in which case $\sum_{ij} x_{ij} = \sum_i \sum_j x_{ij}$.

By symmetry, $\sum_{ij} x_{ij} = \sum_j \sum_i x_{ij}$. Thus the order of summation can be reversed in a nonnegative double series: $\sum_i \sum_j x_{ij} = \sum_j \sum_i x_{ij}$.

A28. *The Weierstrass M-test.*

Theorem. Suppose that $\lim_n x_{nk} = x_k$ for each k and that $|x_{nk}| \leq M_k$, where $\sum_k M_k < \infty$. Then $\sum_k x_k$ and all the $\sum_k x_{nk}$ converge, and $\lim_n \sum_k x_{nk} = \sum_k x_k$.

PROOF. The series of course converge absolutely, since $\sum_k M_k < \infty$. Now $|\sum_k x_{nk} - \sum_k x_k| \leq \sum_{k \leq k_0} |x_{nk} - x_k| + 2\sum_{k > k_0} M_k$. Given ϵ , choose k_0 so that $\sum_{k > k_0} M_k < \epsilon/3$, and then choose n_0 so that $n > n_0$ implies $|x_{nk} - x_k| < \epsilon/3k_0$ for $k \leq k_0$. Then $n > n_0$ implies $|\sum_k x_{nk} - \sum_k x_k| < \epsilon$. ■

A29. Power series. The principal fact needed is this: If $f(x) = \sum_{k=0}^{\infty} a_k x^k$ converges in the range $|x| < r$, then it is differentiable there and

$$(8) \quad f'(x) = \sum_{k=1}^{\infty} k a_k x^{k-1}.$$

For a simple proof, choose r_0 and r_1 so that $|x| < r_0 < r_1 < r$. If $|h| < r_0 - |x|$, so that $|x \pm h| < r_0$, then the mean-value theorem gives (here $0 \leq \theta_h \leq 1$)

$$(9) \quad \left| \frac{(x+h)^k - x^k}{h} - kx^{k-1} \right| = \left| k(x + \theta_h h)^{k-1} - kx^{k-1} \right| \leq 2kr_0^{k-1}.$$

Since $2kr_0^{k-1}/r_1^k$ goes to 0, it is bounded by some M , and if $M_k = |a_k| \cdot Mr_1^k$, then $\sum_k M_k < \infty$ and $|a_k|$ times the left member of (9) is at most M_k for $|h| < r_0 - |x|$. By the M -test [A28] (applied with $h \rightarrow 0$ instead of $n \rightarrow \infty$),

$$\lim_{h \rightarrow 0} \sum_{k=0}^{\infty} a_k \frac{(x+h)^k - x^k}{h} = \sum_{k=0}^{\infty} k a_k x^{k-1}.$$

Hence (8).

Repeated application of (8) gives

$$f^{(j)}(x) = \sum_{k=j}^{\infty} k(k-1) \cdots (k-j+1) a_k x^{k-j}.$$

For $x = 0$, this is $a_j = f^{(j)}(0)/j!$, the formula for the coefficients in a Taylor series. This shows in particular that the values of $f(x)$ for $|x| < r$ determine the coefficients a_k .

A30. Cesàro averages. If $x_n \rightarrow x$, then $n^{-1} \sum_{k=1}^n x_k \rightarrow x$. To prove this, let M bound $|x_k|$, and given ϵ , choose k_0 so that $|x - x_k| < \epsilon/2$ for $k \geq k_0$. If $n > k_0$ and $n > 4k_0 M/\epsilon$, then

$$\left| x - \frac{1}{n} \sum_{k=1}^n x_k \right| \leq \frac{1}{n} \sum_{k=1}^{k_0-1} 2M + \frac{1}{n} \sum_{k=k_0}^n \frac{\epsilon}{2} < \epsilon.$$

A31. Dyadic expansions. Define a mapping T of the unit interval $\Omega = (0, 1]$ into itself by

$$T\omega = \begin{cases} 2\omega & \text{if } 0 < \omega \leq \frac{1}{2}, \\ 2\omega - 1 & \text{if } \frac{1}{2} < \omega \leq 1. \end{cases}$$

Define a function d_1 on Ω by

$$d_1(\omega) = \begin{cases} 0 & \text{if } 0 < \omega \leq \frac{1}{2}, \\ 1 & \text{if } \frac{1}{2} < \omega \leq 1, \end{cases}$$

and let $d_i(\omega) = d_1(T^{i-1}\omega)$. Then

$$(10) \quad \sum_{i=1}^n \frac{d_i(\omega)}{2^i} < \omega \leq \sum_{i=1}^n \frac{d_i(\omega)}{2^i} + \frac{1}{2^n}$$

for all $\omega \in \Omega$ and $n \geq 1$. To verify this for $n = 1$, check the cases $\omega \leq \frac{1}{2}$ and $\omega > \frac{1}{2}$ separately. Suppose that (10) holds for a particular n and for all ω . Replace ω by $T\omega$ in (10) and use the fact that $d_i(T\omega) = d_{i+1}(\omega)$; separate consideration of the cases $\omega \leq \frac{1}{2}$ and $\omega > \frac{1}{2}$ now shows that (10) holds with $n + 1$ in place of n .

Thus (10) holds for all n and ω , and it follows that $\omega = \sum_{i=1}^{\infty} d_i(\omega)/2^i$. This gives the dyadic representation of ω . If $d_i(\omega) = 0$ for all $i > n$, then $\omega = \sum_{i=1}^n d_i(\omega)/2^i$, which contradicts the left-hand inequality in (10). Thus the expansion does not terminate in 0's.

Convex Functions

A32. A function φ on an open interval I (bounded or unbounded) is *convex* if

$$(11) \quad \varphi(tx + (1-t)y) \leq t\varphi(x) + (1-t)\varphi(y)$$

for $x, y \in I$ and $0 \leq t \leq 1$. From this it follows by induction that

$$(12) \quad \varphi\left(\sum_{i=1}^n p_i x_i\right) \leq \sum_{i=1}^n p_i \varphi(x_i)$$

if the x_i lie in I and the p_i are nonnegative and add to 1.

If φ has a continuous, nondecreasing derivative φ' on I , then φ is convex. Indeed, if $a < b < c$, the average of φ' over (a, b) is at most the average over (b, c) :

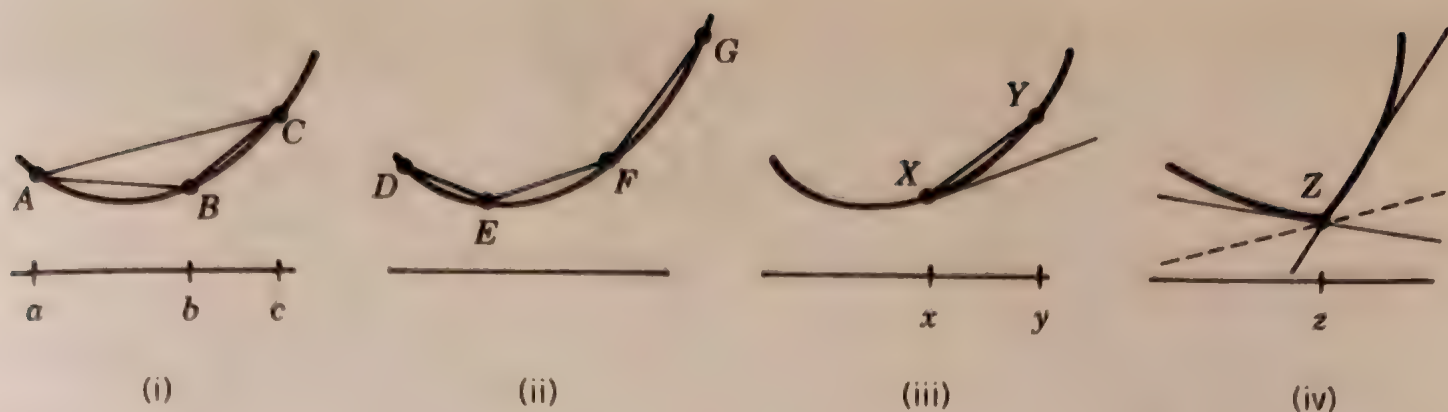
$$\begin{aligned} \frac{\varphi(b) - \varphi(a)}{b - a} &= \frac{1}{b - a} \int_a^b \varphi'(s) ds \leq \varphi'(b) \leq \frac{1}{c - b} \int_b^c \varphi'(s) ds \\ &= \frac{\varphi(c) - \varphi(b)}{c - b}. \end{aligned}$$

The inequality between the extreme terms here reduces to

$$(13) \quad (c - a)\varphi(b) \leq (c - b)\varphi(a) + (b - a)\varphi(c),$$

which is (11) with $x = a$, $y = c$, $t = (c - b)/(c - a)$.

A33. Geometrically, (11) means that the point B in Figure (i) lies on or below the chord AC . But then slope $AB \leq$ slope AC ; algebraically, this is $(\varphi(b) - \varphi(a))/(b - a) \leq (\varphi(c) - \varphi(a))/(c - a)$, which is the same as (13). As B moves to A from the right,



slope AB is thus nonincreasing and hence has a limit. In other words, φ has a right-hand derivative φ^+ . Figure (ii) shows that slope $DE \leq \text{slope } EF \leq \text{slope } FG$. Let E move to D from the right and let G move to F from the right: The right-hand derivative at D is at most that at F , and φ^+ is nondecreasing. Since the slope of XY in Figure (iii) is at least as great as the right-hand derivative at X , the curve to the right of X lies on or above the line through X with slope $\varphi^+(x)$:

$$(14) \quad \varphi(y) \geq \varphi(x) + (y-x)\varphi^+(x), \quad y \geq x.$$

Figure (iii) also makes it clear that φ is right-continuous.

Similarly, φ has a nondecreasing left-hand derivative φ^- and is left-continuous. Since slope $AB \leq \text{slope } BC$ in Figure (i), $\varphi^-(b) \leq \varphi^+(b)$. Since clearly $\varphi^+(b) < \infty$ and $-\infty < \varphi^-(b)$, φ^+ and φ^- are finite. Finally, (14) and its right-sided analogue show that the curve lies on or above each line through Z in Figure (iv) having slope between $\varphi^-(z)$ and $\varphi^+(z)$:

$$(15) \quad \varphi(x) \geq \varphi(z) + m(x-z), \quad \varphi^-(z) \leq m \leq \varphi^+(z).$$

This is a *support line*.

Some Multivariable Calculus

A34. Suppose that U is an open set in R^k and $T: U \rightarrow R^k$ is continuously differentiable; let $D_x = [t_{ij}(x)]$ and $J(x) = \det D_x$ be the Jacobian matrix and determinant, as in Theorem 17.2. Let Q^- be a closed rectangle in U .

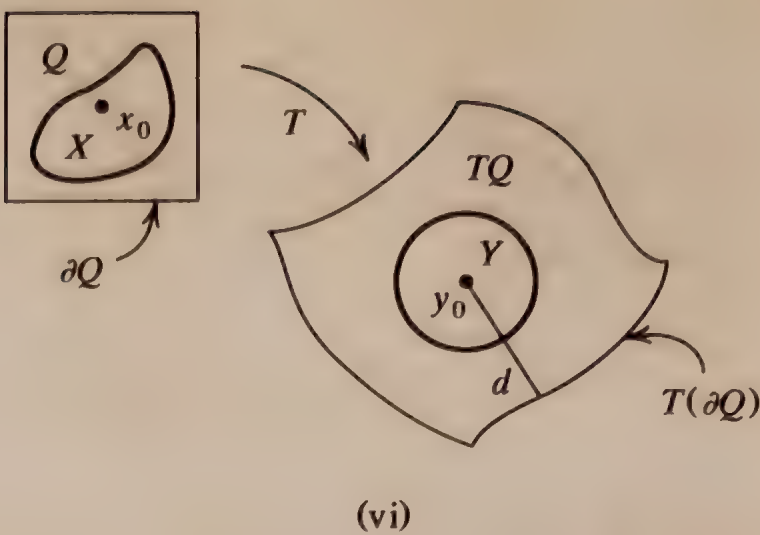
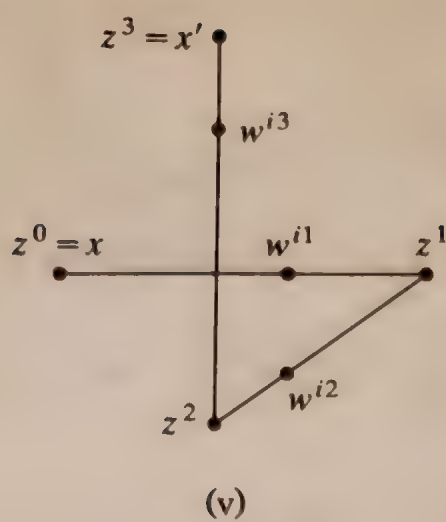
Theorem. If $|t_{ij}(x') - t_{ij}(x)| \leq \alpha$ for $x, x' \in Q^-$ and all i, j , then

$$(16) \quad |Tx' - Tx - D_x(x' - x)| \leq k^2 \alpha |x' - x|, \quad x, x' \in Q.$$

Before proceeding to the proof, note that, since the t_{ij} are continuous, α can be taken to go to 0 as Q contracts to the point x . In other words, (16) implies

$$(17) \quad \lim_{x' \rightarrow x} \frac{|Tx' - Tx - D_x(x' - x)|}{|x' - x|} = 0.$$

This shows that D_x acts as a multivariable derivative. Suppose on the other hand that (17) holds at x for an initially unspecified matrix D_x . Take $x'_j = x_j + h$ and $x'_i = x_i$ for $i \neq j$, and let h go to 0. It follows that the entries of D_x must be the partial derivatives $t_{ij}(x)$: If (17) holds, then D_x must be the Jacobian matrix.



PROOF OF (16). For $j = 0, 1, \dots, k$, let z^j agree with x' in the first j places and with x in the last $k - j$ places. Then $z^0 = x$, $z^k = x'$, and $|z^j - z^{j-1}| = |(z^j - z^{j-1})_j| = |x'_j - x_j|$ (Figure (v)). By the mean-value theorem in one dimension, there is a point w^{ij} on the segment from z^{j-1} to z^j such that $t_i(z^j) - t_i(z^{j-1}) = t_{ij}(w^{ij})(z^j - z^{j-1})_j$. Since

$$(D_x(z^j - z^{j-1}))_i = \sum_l t_{il}(x)(z^j - z^{j-1})_l = t_{ij}(x)(z^j - z^{j-1})_j,$$

it follows that

$$\begin{aligned} &|Tx' - Tx - D_x(x' - x)| \\ &\leq \sum_{ij} |t_i(z^j) - t_i(z^{j-1}) - (D_x(z^j - z^{j-1}))_i| \\ &= \sum_{ij} |t_{ij}(w^{ij}) - t_{ij}(x)| \cdot |(z^j - z^{j-1})_j| \\ &\leq \sum_{ij} \alpha |x'_j - x_j| \leq k^2 \alpha |x' - x|. \end{aligned}$$

A35. The multivariable inverse-function theorem. Let x_0 be a point of the open set U .

Theorem. If $J(x_0) \neq 0$, then there are open sets X and Y , containing x_0 and $y_0 = Tx_0$, respectively, such that T is a one-to-one map from X onto $Y = TX$; further, $T^{-1}: Y \rightarrow X$ is continuously differentiable, and the Jacobian matrix of T^{-1} at y is $D_{T^{-1}}^{-1}y$.

This is a local theorem. It is not assumed, as in Theorem 17.2, that T is one-to-one on U and $J(x)$ never vanishes; but under those additional conditions, TU is open and the inverse point mapping is continuously differentiable. To understand the role of the condition $J(x_0) \neq 0$, consider the case where $k = 1$, $x_0 = 0$, and Tx is x^2 or x^3 .

PROOF. Let Q be a rectangle such that $x_0 \in Q^\circ \subset Q^- \subset U$ and $J(x) \neq 0$ for $x \in Q^-$. As (x, u) ranges over the compact set $Q^- \times [u: |u| = 1]$, $|D_x u|$ is bounded below by some positive β :

(18) $|D_x u| \geq \beta |u| \quad \text{if } x \in Q^-, u \in R^k.$

Making Q smaller will ensure that $|t_{ij}(x) - t_{ij}(x')| \leq \beta/2k^2$ for all x and x' in Q^- and all i, j . Then (16) and (18) give, for $x, x' \in Q^-$,

$$\begin{aligned} |Tx' - Tx| &\geq |D_x(x' - x)| - |Tx' - Tx - D_x(x' - x)| \\ &\geq |D_x(x' - x)| - \frac{1}{2}\beta|x' - x| \geq \frac{1}{2}\beta|x' - x|. \end{aligned}$$

Thus

$$(19) \quad |x' - x| \leq \frac{2}{\beta}|Tx' - Tx| \quad \text{for } x, x' \in Q^-.$$

This shows that T is one-to-one on Q^- .

Since x_0 does not lie in the compact set ∂Q , $\inf_{x \in \partial Q} |Tx - Tx_0| = d > 0$. Let Y be the open ball with center $y_0 = Tx_0$ and radius $d/2$ (Figure (vi)). Fix a y in Y . The problem is to show that $y = Tx$ for some x in Q° , which means finding an x such that $\varphi(x) = |y - Tx|^2 = \sum_i (y_i - t_i(x))^2$ vanishes. By compactness, the minimum of φ on Q^- is achieved there. If $x \in \partial Q$ (and $y \in Y$), then $2|y - y_0| < d \leq |Tx - y_0| \leq |Tx - y| + |y - y_0|$, so that $|y - Tx_0| < |y - Tx|$. Therefore, $\varphi(x_0) < \varphi(x)$ for $x \in \partial Q$, and so the minimum occurs in Q° rather than on ∂Q . At the minimizing point, $\partial\varphi/\partial x_j = -\sum_i 2(y_i - t_i(x))t_{ij}(x) = 0$, and since D_x is nonsingular, it follows that $y = Tx$: Each y in Y is the image under T of some point x in Q° . By (19), this x is unique (although it is possible that $y = Tz$ for some z outside Q).

Let $X = Q^\circ \cap T^{-1}Y$. Then X is open and T is a one-to-one map of X onto Y . Now let T^{-1} denote the inverse point transformation on Y . By (19), T^{-1} is continuous.

To prove differentiability, consider in Y a fixed point y and a variable point y' such that $y' \rightarrow y$ and $y' \neq y$. Let $x = T^{-1}y$ and $x' = T^{-1}y'$; then x' is a function of y' , $x' \rightarrow x$, and $x' \neq x$. Define v by $Tx' - Tx = D_x(x' - x) + v$; then v is a function of x' and hence of y' , and $|v|/|x' - x| \rightarrow 0$ by (17). Apply D_x^{-1} : $D_x^{-1}(Tx' - Tx) = x' - x + D_x^{-1}v$, or $T^{-1}y' - T^{-1}y = D_x^{-1}(y' - y) + D_x^{-1}v$. By (18) and (19),

$$\frac{|T^{-1}y' - T^{-1}y - D_x^{-1}(y' - y)|}{|y' - y|} = \frac{|D_x^{-1}v|}{|x' - x|} \cdot \frac{|x' - x|}{|y' - y|} \leq \frac{|v|/\beta}{|x' - x|} \cdot \frac{2}{\beta}.$$

The right side goes to 0 as $y' \rightarrow y$.

By the remark following (17), the components of D_x^{-1} must be the partial derivatives of the inverse mapping: T^{-1} has Jacobian matrix $D_T^{-1}|_y$ at y . The components of an inverse matrix vary continuously with the components of the original matrix (think for example of the inverse as specified by the cofactors), and so T^{-1} is even continuously differentiable on Y . ■

Continued Fractions

A36. In designing a planetarium, Christian Huygens confronted this problem: Given the ratio x of the periods of two planets, approximate it by the ratio of the periods of two linked gears. If one gear has p teeth and the other has q , then the ratio of their periods is p/q , so that the problem is to approximate the real number x by the rational p/q . Of course x , being empirical, is already rational, but the numerator and denominator may be so large that gears with those numbers of teeth are not practical: in the approximation p/q , both p and q must be of moderate size.

Since the gears play symmetrical roles, there is no more reason to approximate x by $r = p/q$ than there is to approximate $1/x$ by $1/r = q/p$. Suppose, to be definite, that x and r lie to the *left* of 1. Then

$$(20) \quad |x - r| = xr \left| \frac{1}{x} - \frac{1}{r} \right| \leq \left| \frac{1}{x} - \frac{1}{r} \right|,$$

and the inequality is strict unless $x = r$: If $x < 1$, it is better to approximate $1/x$ and then invert the approximation, since that will control *both* errors.

For a numerical illustration, approximate $x = .127$ by rounding it up to $r = .13$; the calculations (to three places) are

$$(21) \quad \begin{aligned} r - x &= .13 - .127 = .003, \\ \frac{1}{x} - \frac{1}{r} &= 7.874 - 7.692 = .182. \end{aligned}$$

The second error is large. So instead, approximate $1/x = 7.874$ by rounding it down to $1/r' = 7.87$. Since $1/7.87 = .1271$ (to four places), the calculations are

$$(22) \quad \begin{aligned} \frac{1}{x} - \frac{1}{r'} &= 7.874 - 7.87 = .004, \\ r' - x &= .1271 - .127 = .0001. \end{aligned}$$

This time, both errors are small, and the error .0001 in the new approximation to x is smaller than the corresponding .003 in (21). It is because x lies to the *left* of 1 that inversion improves the accuracy; see (20).

If this inversion method decreases the error, why not do another inversion in the middle, in finding a rational approximation to $1/x$? It makes no sense to invert $1/x$ itself, since it lies to the right of 1 (and inversion merely leads back to x anyway). But to approximate $1/x = 7.874$ is to approximate the fractional part .874, and here a second inversion will help, for the same reason the first one does. This suggests Huygens's iterative procedure.

In modern notation, the scheme is this. For x (rational or irrational) in $(0, 1)$, let $Tx = \{1/x\}$ and $a_1(x) = [1/x]$ be the fractional and integral parts of $1/x$; and set $T0 = 0$. This defines a mapping of $[0, 1)$ onto itself:

$$(23) \quad Tx = \begin{cases} \left\{ \frac{1}{x} \right\} = \frac{1}{x} - \left[\frac{1}{x} \right] = \frac{1}{x} - a_1(x) & \text{if } 0 < x < 1, \\ 0 & \text{if } x = 0. \end{cases}$$

Then

$$(24) \quad x = \frac{1}{a_1(x) + Tx} \quad \text{if } 0 < x < 1.$$

What (20) says is that replacing Tx on the right in (24) by a good rational approximation to it gives an even better rational approximation to x itself.

To carry this further requires a convenient notation. For positive variables z_i define the *continued fractions*

$$\begin{aligned}\underline{1}z_1 &= \frac{1}{z_1}, & \underline{1}z_1 + \underline{1}z_2 &= \frac{1}{z_1 + \frac{1}{z_2}}, \\ \underline{1}z_1 + \underline{1}z_2 + \underline{1}z_3 &= \frac{1}{z_1 + \frac{1}{z_2 + \frac{1}{z_3}}},\end{aligned}$$

and so on. It is “typographically” clear that

$$(25) \quad \underline{1}z_1 + \cdots + \underline{1}z_n = 1 / \left[z_1 + (\underline{1}z_2 + \cdots + \underline{1}z_n) \right]$$

and

$$(26) \quad \underline{1}z_1 + \cdots + \underline{1}z_n = \underline{1}z_1 + \cdots + \underline{1}z_{n-1} + \underline{1}z_n.$$

For a formal theory, use (25) as a recursive definition and then prove (26) by induction (or vice versa). An infinite continued fraction is defined by

$$\underline{1}z_1 + \underline{1}z_2 + \cdots = \lim_n \underline{1}z_1 + \cdots + \underline{1}z_n,$$

provided the limit exists. A continued fraction is *simple* if the z_i are positive integers.

If $T^{n-1}x > 0$, let $a_n(x) = a_1(T^{n-1}x)$; the $a_n(x)$ are the *partial quotients* of x . If x and Tx are both positive, then (24) applies to each of them:

$$x = \underline{1}a_1(x) + \overline{Tx} = \underline{1}a_1(x) + \underline{1}a_2(x) + \overline{T^2x}.$$

If none of the iterates $x, Tx, \dots, T^{n-1}x$ vanishes, then it follows by induction (use (26)) that

$$(27) \quad x = \underline{1}a_1(x) + \cdots + \underline{1}a_{n-1}(x) + \underline{1}a_n(x) + \overline{T^n x}.$$

This is an extension of (24), and the idea following (24) extends as well: a good rational approximation to $T^n x$ in (27) gives a still better rational approximation to x . Even if $T^n x$ is approximated very crudely by 0, there results a sharp approximation

$$(28) \quad x \approx \underline{1}a_1(x) + \cdots + \underline{1}a_n(x)$$

to x itself. The right side here is the n th *convergent* to x , and it goes very rapidly to x ; see Section 24.

By the definition (23), x and Tx are both rational or both irrational. For an irrational x , therefore, $T^n x$ remains forever among the irrationals, and (27) holds for all n . If x is rational, on the other hand, $T^n x$ remains forever among the rationals, and in fact, as the following argument shows, $T^n x$ eventually hits 0 and stays there.

Suppose that x is a rational in $(0, 1)$: $x = d_1/d_0$, where $0 < d_1 < d_0$. If $Tx > 0$, then $Tx = \{d_0/d_1\} = d_2/d_1$, where $0 < d_2 < d_1$ because $0 < Tx < 1$. (If d_1/d_0 is irreducible, so is d_2/d_1 .) If $T^2x > 0$, the argument can be repeated:

$$(29) \quad x = \frac{d_1}{d_0}, \quad Tx = \frac{d_2}{d_1}, \quad T^2x = \frac{d_3}{d_2}, \quad d_0 > d_1 > d_2 > d_3 > 0.$$

And so on. Since the d_n decrease as long as the $T^n x$ remain positive, $T^n x$ must vanish for some n , and then $T^m x = 0$ for $m \geq n$. If n_x is the smallest integer for which $T^{n_x} x = 0$ ($n_x \geq 1$ if $x > 0$), then by (27),

$$(30) \quad x = \underline{1} \overline{a_1(x)} + \cdots + \underline{1} \overline{a_{n_x}(x)}.$$

Thus each positive rational has a representation as a finite simple continued fraction. If $0 < x < 1$ and $Tx = 0$, then $1 > x = 1/a_1(x)$, so that $a_1(x) \geq 2$. Applied to $T^{n_x-1}x$, this shows that the $a_{n_x}(x)$ in (30) must be at least 2.

Section 24 requires a uniqueness result. Suppose that

$$(31) \quad x = \underline{1} \overline{a_1} + \cdots + \underline{1} \overline{a_{n-1}} + \underline{1} \overline{a_n + t},$$

where the a_i are positive integers and

$$(32) \quad 0 < x < 1, \quad 0 \leq t < 1, \quad a_n + t > 1.$$

The last condition rules out $a_n = 1$ and $t = 0$ (which in the case $n = 1$ is also ruled out by $x < 1$). It follows from (31) and (32) that

$$(33) \quad a_1(x) = a_1, \dots, a_n(x) = a_n, \quad T^n x = t.$$

The case $n = 1$ being easy, suppose the implication holds for $n - 1$, where $n \geq 2$. Since $0 < 1/(a_n + t) < 1$, the induction hypothesis (use (26)) gives $a_k(x) = a_k$ for $k < n$ and $T^{n-1}x = 1/(a_n + t)$. Now apply the case $n = 1$ to $T^{n-1}x$. (If $a_n = 1$ and $t = 0$, then $a_k(x) = a_k$ for $k \leq n - 2$, $a_{n-1}(x) = a_{n-1} + 1$, and $T^{n-1}x = 0$.)

Consider now the infinite case. Assume that

$$(34) \quad x = \underline{1} \overline{a_1} + \underline{1} \overline{a_2} + \cdots,$$

converges, where the a_n are positive integers. Then

$$(35) \quad a_n(x) = a_n, \quad T^n x = \underline{1} \overline{a_{n+1}} + \underline{1} \overline{a_{n+2}} + \cdots, \quad n \geq 1.$$

To prove this, let $n \rightarrow \infty$ in (25): the continued fraction $t = \underline{1} \overline{a_2} + \underline{1} \overline{a_3} + \cdots$ converges and $x = 1/(a_1 + t)$. It follows by induction (use (26)) that

$$(36) \quad \underline{1} \overline{a_1} > \underline{1} \overline{a_1} + \cdots + \underline{1} \overline{a_n} \geq \underline{1} \overline{a_1} + \underline{1} \overline{a_2}, \quad n \geq 2.$$

Hence $0 < x < 1$, and the same must be true of t . Therefore, a_1 and t are the integer and fractional parts of $1/x$, which proves (35) for $n = 1$. Apply the same argument to

Tx , and continue. The x defined by (34) is irrational: otherwise, $T^n x = 0$ for some n , which contradicts (35) and (36).

Thus the value of an infinite simple continued fraction uniquely determines the partial quotients. The same is almost true of finite simple continued fractions. Since (31) and (32) imply (33), it follows that if x is given by (30), then any continued fraction of n_x terms that represents x must indeed match (30) term for term. But, for example, $\frac{1}{\sqrt{3}} + \frac{1}{\sqrt{5}} = \frac{1}{\sqrt{3}} + \frac{1}{\sqrt{4}} + \frac{1}{\sqrt{1}}$. This is always possible: replace $a_{n_x}(x)$ in (30) (where $a_{n_x}(x) \geq 2$) by $a_{n_x}(x) - 1 + \frac{1}{\sqrt{1}}$. Apart from this ambiguity, the representation is unique—and the representation (30) that results from repeated application of T to a rational x never ends with a partial quotient of 1.[†]

[†]See ROCKETT & SZÜSZ for more on continued fractions.

Notes on the Problems

These notes consist of hints, solutions, and references to the literature. As a rule a solution is complete in proportion to the frequency with which it is needed for the solution of subsequent problems.

Section 1

1.1. (a) Each point of the discrete space lies in one of the four sets $A_1 \cap A_2$, $A_1^c \cap A_2$, $A_1 \cap A_2^c$, $A_1^c \cap A_2^c$ and hence would have probability at most 2^{-2} ; continue.

(b) If, for each i , B_i is A_i or A_i^c , then $B_1 \cap \cdots \cap B_n$ has probability at most $\prod_{i=1}^n (1 - \alpha_i) \leq \exp[-\sum_{i=1}^n \alpha_i]$.

1.3. (b) Suppose A is trifling and let A^- be its closure. Given ϵ choose intervals $(a_k, b_k]$, $k = 1, \dots, n$, such that $A \subset \bigcup_{k=1}^n (a_k, b_k]$ and $\sum_{k=1}^n (b_k - a_k) < \epsilon/2$. If $x_k = a_k - \epsilon/2n$, then $A^- \subset \bigcup_{k=1}^n (x_k, b_k]$ and $\sum_{k=1}^n (b_k - x_k) < \epsilon$.

For the other parts of the problem, consider the set of rationals in $(0, 1)$.

1.4. (a) Cover $A_r(i)$ by $(r - 1)^n$ intervals of length r^{-n} .

(c) Go to the base r^k . Identify the digits in the base r with the keys of the typewriter. The monkey is certain eventually to reproduce the eleventh edition of the *Britannica* and even, unhappily, the fifteenth.

1.5. (a) The set $A_3(1)$ is itself uncountable, since a point in it is specified by a sequence of 0's and 2's (excluding the countably many that end in 0's).

(b) For sequences u_1, \dots, u_n of 0's, 1's, and 2's, let M_{u_1, \dots, u_n} consist of the points in $(0, 1]$ whose nonterminating base-3 expansions start out with those digits. Then $A_3(1) = (0, 1] - \bigcup M_{u_1, \dots, u_n}$, where the union extends over $n \geq 1$ and the sequences u_1, \dots, u_n containing at least one 1. The set described in part (b) is $[0, 1] - \bigcup M_{u_1, \dots, u_n}^o$, where the union is as before, and this is the closure of $A_3(1)$.

From this representation of C , it is not hard to deduce that it can be defined as the set of points in $[0, 1]$ that can be written in base 3 without any 1's if terminating expansions are also allowed. For example, C contains $\frac{2}{3} = .1222 \cdots = .2000 \cdots$ because it is possible to avoid 1 in the expansion.

(c) Given ϵ and an ω in C , choose ω' in $A_3(1)$ within $\epsilon/2$ of ω ; now define ω'' by changing from 2 to 0 some digit of ω' far enough out that ω'' differs from ω' by at most $\epsilon/2$.

- 1.7. The interchange of limit and integral is justified because the series $\sum_k r_k(\omega)2^{-k}$ converges uniformly in ω (integration to the limit is studied systematically in Section 16). There is a direct derivation of (1.40): let $n \rightarrow \infty$ in $\sin t = 2^n \sin 2^{-n} t \cdot \prod_{k=1}^n \cos 2^{-k} t$, which follows by induction from the half-angle formula.
- 1.10. (a) Given m and a subinterval $(a, b]$ of $(0, 1]$, choose a dyadic interval I in $(a, b]$, and then choose in I a dyadic interval J of order $n > m$ such that $|n^{-1} s_n(\omega)| > \frac{1}{2}$ for $\omega \in J$. This is possible because to specify J is to specify the first n dyadic digits of the points in J ; choose the first digits in such a way that $J \subset I$ and take the following ones to be 1, with n so large that $n^{-1} s_n(\omega)$ is near 1 for $\omega \in J$.
- (b) A countable union of sets of the first category is also of the first category; $(0, 1] = N \cup N^c$ would be of the first category if N^c were. For Baire's theorem, see ROYDEN, p. 139.
- 1.11. (a) If $x = p_0/q_0 \neq p/q$, then
- $$\left| x - \frac{p}{q} \right| = \frac{|p_0 q - q_0 p|}{q_0 q} \geq \frac{1}{q_0 q}.$$
- (c) The rational $\sum_{k=1}^n 1/2^{\alpha(k)}$ has denominator $2^{\alpha(n)}$ and approximates x to within $2/2^{\alpha(n+1)}$.

Section 2

- 2.3. (b) Let Ω consist of four points, and let \mathcal{F} consist of the empty set, Ω itself, and all six of the two-point sets.
- 2.4. (b) For example, take Ω to consist of the integers, and let \mathcal{F}_n be the σ -field generated by the singletons $\{k\}$ with $k \leq n$. As a matter of fact, any example in which \mathcal{F}_n is a proper subclass of \mathcal{F}_{n+1} for all n will do, because it can be shown that in this case $\bigcup_n \mathcal{F}_n$ necessarily fails to be a σ -field; see A. Broughton and B. W. Huff: A comment on unions of sigma-fields, *Amer. Math. Monthly*, **84** (1977), 553–554.
- 2.5. (b) The class in question is certainly contained in $f(\mathcal{A})$ and is easily seen to be closer under the formation of finite intersections. But $(\bigcup_{i=1}^m \bigcap_{j=1}^{n_i} A_{ij})^c = \bigcap_{i=1}^m \bigcup_{j=1}^{n_i} A_{ij}^c$, and $\bigcup_{j=1}^{n_i} A_{ij}^c = \bigcup_{j=1}^{n_i} [A_{ij}^c \cap \bigcap_{k=1}^{j-1} A_{ik}]$ has the required form.
- 2.8. If \mathcal{K} is the smallest class over \mathcal{A} closed under the formation of countable unions and intersections, clearly $\mathcal{K} \subset \sigma(\mathcal{A})$. To prove the reverse inclusion, first show that the class of A such that $A^c \in \mathcal{K}$ is closed under the formation of countable unions and intersections and contains \mathcal{A} and hence contains \mathcal{K} .
- 2.9. Note that $\bigcup_n B_n \in \sigma(\bigcup_n \mathcal{A}_{B_n})$.
- 2.10. (a) Show that the class of A for which $I_A(\omega) = I_A(\omega')$ is a σ -field. See Example 4.8.
- 2.11. (b) Suppose that \mathcal{F} is the σ -field of the countable and the cocountable sets in Ω . Suppose that \mathcal{F} is countably generated and Ω is uncountable. Show that \mathcal{F}

is generated by a countable class of singletons; if Ω_0 is the union of these, then \mathcal{F} must consist of the sets B and $B \cup \Omega_0^c$ with $B \subset \Omega_0$, and these do not include the singletons in Ω_0^c , which is uncountable because Ω is.

(c) Let \mathcal{F}_2 consist of the Borel sets in $\Omega = (0, 1]$, and let \mathcal{F}_1 consist of the countable and the cocountable sets there.

2.12. Suppose that A_1, A_2, \dots is an infinite sequence of distinct sets in a σ -field \mathcal{F} , and let \mathcal{G} consist of the nonempty sets of the form $\bigcap_{n=1}^{\infty} B_n$, where $B_n = A_n$ or $B_n = A_n^c$, $n = 1, 2, \dots$. Each A_n is the union of the \mathcal{G} -sets it contains, and since the A_n are distinct, \mathcal{G} must be infinite. But there are uncountably many distinct countable unions of \mathcal{G} -sets, and they all lie in \mathcal{F} .

2.18. For this and the subsequent problems on applications of probability theory to arithmetic, the only number theory required is the fundamental theorem of arithmetic and its immediate consequences. The other problems on stochastic arithmetic are 4.15, 4.16, 5.19, 5.20, 6.16, 18.17, 25.15, 30.9, 30.10, 30.11, and 30.12. See also Theorem 30.3.

(b) Let A consist of the even integers, let $C_k = [m: \nu_k < m \leq \nu_{k+1}]$, and let B consist of the even integers in $C_1 \cup C_3 \cup \dots$ together with the odd integers in $C_2 \cup C_4 \cup \dots$; take ν_k to increase very rapidly with k and consider $A \cap B$.

(c) If c is the least common multiple of a and b , then $M_a \cap M_b = M_c$. From $M_a \in \mathcal{D}$ conclude in succession that $M_a \cap M_b \in \mathcal{D}$, $M_{a_1} \cap \dots \cap M_{a_j} \cap M_{b_1}^c \cap \dots \cap M_{b_k}^c \in \mathcal{D}$, $f(\mathcal{M}) \subset \mathcal{D}$. By the same sequence of steps, show how D on \mathcal{M} determines D on $f(\mathcal{M})$.

(d) If $B_l = M_a - \bigcup_{p \leq l} M_{ap}$, then $a \in B_l$ and (the inclusion-exclusion formula requires only finite additivity)

$$\begin{aligned} D(B_l) &= \frac{1}{a} - \sum_{p \leq l} \frac{1}{ap} + \sum_{p < q \leq l} \frac{1}{apq} - \dots \\ &= \frac{1}{a} \prod_{p \leq l} \left(1 - \frac{1}{p}\right) \leq \frac{1}{a} \exp\left(-\sum_{p \leq l} \frac{1}{p}\right) \rightarrow 0. \end{aligned}$$

Choose l_a so that, if $C_a = B_{l_a}$, then $D(C_a) < 2^{-a-1}$. If D were a probability measure on $f(\mathcal{M})$, $D(\Omega) \leq \frac{1}{2}$ would follow. See Problem 4.16 for a different approach.

2.19. (a) Apply the intermediate-value theorem to the function $f(x) = \lambda(A \cap (0, x])$. Note that this even proves part (c) for λ (under the assumption that λ exists).

(b) If $0 < P(B) < P(A)$, then either $0 < P(B) \leq \frac{1}{2}P(A)$ or $0 < P(B - A) \leq \frac{1}{2}P(A)$. Continue.

(c) If $P(\bigcup_k H_k) < x$, choose C so that $C \subset A - \bigcup_k H_k$ and $0 < P(C) < x - P(\bigcup_k H_k)$. If $n^{-1} < P(C)$, then $P(\bigcup_{k < n} H_k) + h_n < P(\bigcup_{k < n} H_k) + P(H_n) + P(C) \leq P(\bigcup_{k < n} H_k) + h_n$.

2.21. (c) If \mathcal{J}_{n-1} were a σ -field, $\mathcal{B} \subset \mathcal{J}_{n-1}$ would follow.

2.22. Use the fact that, if $\alpha_1, \alpha_2, \dots$ is a sequence of ordinals satisfying $\alpha_n < \Omega$, then there exists an ordinal α such that $\alpha < \Omega$ and $\alpha_n < \alpha$ for all n .

- 2.23. Suppose that $B_j \in \bigcup_{\beta < \alpha} \mathcal{I}_\beta$, $j = 1, 2, \dots$. Choose odd integers n_j in such a way that $B_j \in \mathcal{I}_{\beta_\alpha(n_j)}$ and the n_j are all distinct; choose \mathcal{I}_0 -sets such that

$$\Phi_{\beta_\alpha(n_j)}(C_{m_{n_j1}}, C_{m_{n_j2}}, \dots) = B_j;$$

for n not of the form n_j , choose \mathcal{I}_0 -sets for which $\Phi_{\beta_\alpha(n)}(C_{m_{n1}}, C_{m_{n2}}, \dots)$ is \emptyset or $(0, 1]$ as n is odd or even. Then $\bigcup_{j=1}^{\infty} B_j = \Phi_\alpha(C_1, C_2, \dots)$. Similarly, $B^c = \Phi_\alpha(C_1, C_2, \dots)$ for \mathcal{I}_0 -sets C_n if $B \in \bigcup_{\beta < \alpha} \mathcal{I}_\beta$. The rest of the proof is essentially the same as before.

Section 3

- 3.1. (a) The finite additivity of P is used in the proof that $\mathcal{F}_0 \subset \mathcal{M}$ and again (via monotonicity; see (2.5)) in the proof of (3.7). The countable additivity of P is used (via countable subadditivity; see Theorem 2.1) in the proof of (3.7).
 (b) For a specific example consider $\bigcup_n (2^{-1} + n^{-1}, 1]$ in connection with Problem 2.15. But an example is provided by *every* P that is finitely but not countably additive: If P is finitely additive on a field \mathcal{F}_0 and A_n are disjoint \mathcal{F}_0 -sets whose union A also lies in \mathcal{F}_0 , then monotonicity (which requires finite additivity only) gives $\sum_{k \leq n} P(A_k) = P(\bigcup_{k \leq n} A_k) \leq P(A)$ and hence $\sum_k P(A_k) \leq P(A)$. Countable subadditivity will ensure that there is equality here.
 (c) The proof of (3.7) involves the countable subadditivity of P on \mathcal{F}_0 , which is only assumed to be a field (that being the whole point of the theorem).
- 3.2. (a) Given ϵ , choose \mathcal{F}_0 -sets A_n such that $A \subset \bigcup_n A_n$ and $\sum P(A_n) < P^*(A) + \epsilon$; if $B = \bigcup_n A_n$, then $A \subset B$, $B \in \mathcal{F}$, and $P(B) < P^*(A) + \epsilon$; hence the right side of (3.9) is at most $P^*(A)$. On the other hand, $A \subset B$ and $B \in \mathcal{F}$ imply $P^*(A) \leq P^*(B) = P(B)$. Hence (3.9). If $A \subset B_k$, $B_k \in \mathcal{F}$, $P(B_k) < P^*(A) + k^{-1}$, and $B = \bigcap_k B_k$, then $A \subset B$, $B \in \mathcal{F}$, and $P^*(A) = P(B)$. For (3.10), argue by complementation.
 (b) Suppose that $P_*(A) = P^*(A)$ and chose \mathcal{F} -sets A_1 and A_2 in such a way that $A_1 \subset A \subset A_2$ and $P(A_1) = P(A_2)$. Given E , choose an \mathcal{F} -set B in such a way that $E \subset B$ and $P^*(E) = P(B)$. Then $P^*(A \cap E) + P^*(A^c \cap E) \leq P(A_2 \cap B) + P(A_1^c \cap B)$. Now use (2.7) to bound the last sum by $P(B) + P(A_2 - A_1) = P^*(E)$.
- 3.3. First note the general fact that P^* agrees with P on \mathcal{F}_0 if and *only if* P is countably additive there, a condition not satisfied in parts (b) and (e). Using Problem 3.2 simplifies the analysis of P^* and $\mathcal{M}(P^*)$ in the other parts.
 Note in parts (b) and (e) that, if P^* and P_* are defined by (3.1) and (3.2), then, since $P^*(A) = 0$ for all A , (3.4) holds for all A and (3.3) holds for no A . Countable additivity thus plays an essential role in Problem 3.2.
- 3.6 (c) Split E^c by A : $P_\circ(E) = 1 - P^\circ(E^c) = 1 - P^\circ(A \cap E^c) - P^\circ(A^c \cap E^c) = 1 - P^\circ(A \cap E^c) - P(A^c) = P(A) - P^\circ(A - E)$.
- 3.7. (b) Apply (3.13): For $A \in \mathcal{F}_0$, $Q(A) = P^\circ(H \cap A) + P_\circ(H^c \cap A) = P^\circ(H \cap A) + P(A) - P^\circ(A - (H^c \cap A)) = P(A)$.
 (c) If A_1 and A_2 are disjoint \mathcal{F}_0 -sets, then by (3.12),

$$P^\circ(H \cap (A_1 \cup A_2)) = P^\circ(H \cap A_1) + P^\circ(H \cap A_2).$$

Apply (3.13) to the three terms in this equation, successively using $A_1 \cup A_2$, A_1 , and A_2 for A :

$$P_o(H^c \cap (A_1 \cup A_2)) = P_o(H^c \cap A_1) + P_o(H^c \cap A_2).$$

But for these two equations to hold it is enough that $H \cap A_1 \cap A_2 = \emptyset$ in the first case and $H^c \cap A_1 \cap A_2 = \emptyset$ in the second (replacing A_1 by $A_1 \cap A_2^c$ changes nothing).

- 3.8.** By using Banach limits (BANACH, p. 34) one can similarly prove that density D on the class \mathcal{D} (Problem 2.18) extends to a finitely additive probability on the class of all subsets of $\Omega = \{1, 2, \dots\}$.
- 3.14.** The argument is based on cardinality. Since the Cantor set C has Lebesgue measure 0, 2^C is contained in the class \mathcal{L} of Lebesgue sets in $(0, 1]$. But C is uncountable: $\text{card } \mathcal{B} = \text{card}(0, 1] < \text{card } 2^C \leq \text{card } \mathcal{L}$.
- 3.18.** (a) Since the $A \oplus r$ are disjoint Borel sets, $\sum_r \lambda(A \oplus r) \leq 1$, and so the common value $\lambda(A)$ of the $\lambda(A \oplus r)$ must be 0. Similarly, if A is a Borel set contained in some $H \oplus r$, then $\lambda(A) = 0$.
 (b) If the $E \cap (H \oplus r)$ are all Borel sets, they all have Lebesgue measure 0, and so E is a Borel set of Lebesgue measure 0.
- 3.19.** (b) Given $A_1, B_1, \dots, A_{n-1}, B_{n-1}$, note that their union C_n is nowhere dense, so that I_n contains an interval J_n disjoint from C_n . Choose in J_n disjoint, nowhere dense sets A_n and B_n of positive measure.
 (c) Note that A and B_n are disjoint and that $A_n \cup B_n \subset G$.
- 3.20.** (a) If I_n are disjoint open intervals with union G , then $b^{-1}\lambda(A) > \sum_n \lambda(I_n) \geq \sum_n b^{-1}\lambda(A \cap I_n) \geq b^{-1}\lambda(A)$.

Section 4

- 4.1.** Let r be the quantity on the right in (4.30), assumed finite. Suppose that $x < r$; then $x < \bigvee_{k=n}^{\infty} x_k$ for $n \geq 1$ and hence $x < x_k$ for some $k \geq n$: $x < x_n$ i.o. Suppose that $x < x_n$ i.o.; then $x < \bigvee_{k=n}^{\infty} x_k$ for $n \geq 1$: $x \leq r$. It follows that $r = \sup\{x: x < x_n \text{ i.o.}\}$, which is easily seen to be the supremum of the limit points of the sequence. The argument for (4.31) is similar.
- 4.10.** The class \mathcal{F} is the σ -field generated by $\mathcal{G} \cup \{H\}$ (Problem 2.7(a)). If $(H \cap G_1) \cup (H^c \cap G_2) = (H \cap G'_1) \cup (H^c \cap G'_2)$, then $G_1 \Delta G'_1 \subset H^c$ and $G_2 \Delta G'_2 \subset H$; consistency now follows because $\lambda_*(H) = \lambda_*(H^c) = 0$. If $A_n = (H \cap G_1^{(n)}) \cup (H^c \cap G_2^{(n)})$ are disjoint, then $G_1^{(m)} \cap G_1^{(n)} \subset H^c$ and $G_2^{(m)} \cap G_2^{(n)} \subset H$ for $m \neq n$, and therefore (see Problem 2.17) $P(\bigcup_n A_n) = \frac{1}{2}\lambda(\bigcup_n G_1^{(n)}) + \frac{1}{2}\lambda(\bigcup_n G_2^{(n)}) = \sum_n (\frac{1}{2}\lambda(G_1^{(n)}) + \frac{1}{2}\lambda(G_2^{(n)})) = \sum_n P(A_n)$. The intervals with rational endpoints generate \mathcal{G} .
- 4.14.** Show as in Problem 1.1(b) that the maximum of $P(B_1 \cap \dots \cap B_n)$, where B_i is A_i or A_i^c , goes to 0. Let $A_x = [\omega: \sum_n I_{A_n}(\omega) 2^{-n} \leq x]$, show that $P(A \cap A_x)$ is continuous in x , and proceed as in Problem 2.19(a).

4.15. Calculate $D(F_l)$ by (2.36) and the inclusion–exclusion formula, and estimate $P_n(F_l - F)$ by subadditivity; now use $0 \leq P_n(F_l) - P_n(F) = P_n(F_l - F)$. For the calculation of the infinite product, see HARDY & WRIGHT, p. 246.

Section 5

- 5.5. (a) If $m = 0$, $\alpha \geq 0$, and $x > 0$, then $P[X \geq \alpha] \leq P[(X + x)^2 \geq (\alpha + x)^2] \leq E[(X + x)^2]/(\alpha + x)^2 = (\sigma^2 + x^2)/(\alpha + x)^2$; minimize over x .
- 5.8. (b) It is enough to prove that $\varphi(t) = f(t(x', y') + (1 - t)(x, y))$ is convex in t ($0 \leq t \leq 1$) for (x, y) and (x', y') in C . If $\alpha = x' - x$ and $\beta = y' - y$, then (if $f_{11} > 0$)

$$\begin{aligned}\varphi'' &= f_{11}\alpha^2 + 2f_{12}\alpha\beta + f_{22}\beta^2 \\ &= \frac{1}{f_{11}}(f_{11}\alpha + f_{12}\beta)^2 + \frac{1}{f_{11}}(f_{11}f_{22} - f_{12}^2)\beta^2 \geq 0.\end{aligned}$$

Examples like $f(x, y) = y^2 - 2xy$ show that convexity in each variable separately does not imply convexity.

- 5.9. Check (5.39) for $f(x, y) = -x^{1/p}y^{1/q}$.
- 5.10. Check (5.39) for $f(x, y) = -(x^{1/p} + y^{1/p})^p$.
- 5.19. For (5.43) use (2.36) and the fundamental theorem of arithmetic: since the p_i are distinct, the $p_i^{k_i}$ individually divide m if and only if their product does. For (5.44) use inclusion–exclusion. For (5.47), use (5.29) (see Problem 5.12)).
- 5.20. (a) By (5.47), $E_n[\alpha_p] \leq \sum_{k=1}^\infty p^{-k} \leq 2/p$. And, of course, $n^{-1} \log n! = E_n[\log] = \sum_p E_n[\alpha_p] \log p$.
- (b) Use (5.48) and the fact that $E_n[\alpha_p - \delta_p] \leq \sum_{k=2}^\infty p^{-k}$.
- (c) By (5.49),

$$\begin{aligned}\sum_{n < p \leq 2n} \log p &= \sum_{n < p \leq 2n} \left(\left\lfloor \frac{2n}{p} \right\rfloor - 2 \left\lfloor \frac{n}{p} \right\rfloor \right) \log p \\ &\leq 2n(E_{2n}[\log^*] - E_n[\log^*]) = O(n).\end{aligned}$$

Deduce (5.50) by splitting the range of summation by successive powers of 2.

(d) If K bounds the $O(1)$ terms in (5.51), then

$$\sum_{p \leq x} \log p \geq \theta x \sum_{\theta x < p \leq x} p^{-1} \log p \geq \theta x (\log \theta^{-1} - 2K).$$

(e) For (5.53) use

$$\begin{aligned}\sum_{p \leq x} \frac{\log p}{\log x} &\leq \pi(x) \leq \sum_{p \leq x^{1/2}} 1 + \sum_{x^{1/2} < p \leq x} \frac{\log p}{\log x^{1/2}} \\ &\leq x^{1/2} + \frac{2}{\log x} \sum_{p \leq x} \log p.\end{aligned}$$

By (5.53), $\pi(x) \geq x^{1/2}$ for large x , and hence $\log \pi(x) \asymp \log x$ and $\pi(x) \asymp x/\log \pi(x)$. Apply this with $x = p_r$ and note that $\pi(p_r) = r$.

Section 6

6.3. Since for given values of $X_{n1}(\omega), \dots, X_{n,k-1}(\omega)$ there are for $X_n(\omega)$ the k possible values $0, 1, \dots, k-1$, the number of values of $(X_{n1}(\omega), \dots, X_{nn}(\omega))$ is $n!$. Therefore, the map $\omega \rightarrow (X_{n1}(\omega), \dots, X_{nn}(\omega))$ is one-to-one, and the $X_{nk}(\omega)$ determine ω . It follows that if $0 \leq x_i < i$ for $1 \leq i \leq k$, then the number of permutations ω satisfying $X_{ni}(\omega) = x_i$, $1 \leq i \leq k$, is just $(k+1)(k+2) \cdots n$, so that $P[X_{ni} = x_i, 1 \leq i \leq k] = 1/k!$. It now follows by induction on k that X_{n1}, \dots, X_{nk} are independent and $P[X_{nk} = x] = k^{-1}$ ($0 \leq x < k$).

Now calculate

$$\begin{aligned} E[X_{nk}] &= \frac{k-1}{2}, \\ E[S_n] &= \frac{0+1+\cdots+(n-1)}{2} = \frac{n(n-1)}{4} \sim \frac{n^2}{4}, \\ \text{Var}[X_{nk}] &= \frac{0^2+1^2+\cdots+(k-1)^2}{k} - \left(\frac{k-1}{2}\right)^2 = \frac{k^2-1}{12}, \\ \text{Var}[S_n] &= \frac{1}{12} \sum_{k=1}^n (k^2-1) = \frac{2n^3+3n^2-5n}{72} \sim \frac{n^3}{36}. \end{aligned}$$

Apply Chebyshev's inequality.

6.7. (a) If $k^2 \leq n < (k+1)^2$, let $a_n = k^2$; if M bounds the $|x_n|$, then

$$\left| \frac{1}{n} s_n - \frac{1}{a_n} s_{a_n} \right| \leq \left| \frac{1}{n} - \frac{1}{a_n} \right| \cdot nM + \frac{1}{a_n} (n - a_n)M = 2M \frac{n - a_n}{a_n} \rightarrow 0.$$

6.16. From (5.53) and (5.54) it follows that $a_n = \sum_p n^{-1} \lfloor n/p \rfloor \rightarrow \infty$. The left side of (6.8) is

$$\frac{1}{n} \left\lfloor \frac{n}{pq} \right\rfloor - \frac{1}{n} \left\lfloor \frac{n}{p} \right\rfloor \frac{1}{n} \left\lfloor \frac{n}{q} \right\rfloor \leq \frac{1}{pq} - \left(\frac{1}{p} - \frac{1}{n} \right) \left(\frac{1}{q} - \frac{1}{n} \right) \leq \frac{1}{np} + \frac{1}{nq}.$$

Section 7

7.3. If one grants that there are only countably many effective rules, the result is an immediate consequence of the mathematics of this and the preceding sections: C is a countable intersection of \mathcal{F} -sets of measure 1. The argument proves in particular the nontrivial fact that collectives exist.

7.7. If $n \leq \tau$, then $W_n = W_{n-1} - X_{n-1} = W_1 - S_{n-1}$, and τ is the smallest n for which $S_{n-1} = W_1$. Use (7.8) for the question of whether the game terminates. Now

$$F_\tau = F_0 + \sum_{k=1}^{\tau-1} (W_1 - S_{k-1}) X_k = F_0 + W_1 S_{\tau-1} - \frac{1}{2} (S_{\tau-1}^2 - (\tau-1)).$$

- 7.8. Let x_1, \dots, x_i be the initial pattern and put $\Sigma_0 = x_1 + \dots + x_i$. Define $\Sigma_n = \Sigma_{n-1} - W_n X_n$, $L_0 = k$, and $L_n = L_{n-1} - (3X_n + 1)/2$. Then τ is the smallest n such that $L_n \leq 0$, and τ is by the strong law finite with probability 1 if $E[3X_n + 1] = 6(p - \frac{1}{3}) > 0$. For $n \leq \tau$, Σ_n is the sum of the pattern used to determine W_{n+1} . Since $F_n - F_{n-1} = \Sigma_{n-1} - \Sigma_n$, it follows that $F_n = F_0 + \Sigma_0 - \Sigma_n$ and $F_\tau = F_0 + \Sigma_0$.
- 7.9. Observe that $E[F_n - F_\tau] = E[\sum_{k=1}^n X_k I_{[\tau < k]}] = \sum_{k=1}^n E[X_k] P[\tau < k]$.

Section 8

- 8.8. (b) With probability 1 the population either dies out or goes to infinity. If, for example, $p_{k0} = 1 - p_{k,k+1} = 1/k^2$, then extinction and explosion each have positive probability.
- 8.9. To prove that $x_i \equiv 0$ is the only possibility in the persistent case, use Problem 8.5, or else argue directly: If $x_i = \sum_{j \neq i_0} p_{ij} x_j$, $i \neq i_0$, and K bounds the $|x_i|$, then $x_i = \sum p_{ij_1} \dots p_{j_{n-1}j_n} x_{j_n}$, where the sum is over j_1, \dots, j_{n-1} distinct from i_0 , and hence $|x_i| \leq K P_i[X_k \neq i_0, k \leq n] \rightarrow 0$.
- 8.13. Let P be the set of i for which $\pi_i > 0$, let N be the set of i for which $\pi_i \leq 0$, and suppose that P and N are both nonempty. For $i_0 \in P$ and $j_0 \in N$ choose n so that $p_{i_0 j_0}^{(n)} > 0$. Then

$$\begin{aligned} 0 < \sum_{j \in N} \sum_{i \in P} \pi_i p_{ij}^{(n)} &= \sum_{j \in N} \pi_j - \sum_{j \in N} \sum_{i \in N} \pi_i p_{ij}^{(n)} \\ &= \sum_{i \in N} \pi_i \sum_{j \in P} p_{ij}^{(n)} \leq 0. \end{aligned}$$

Transfer from N to P any i for which $\pi_i = 0$ and use a similar argument.

- 8.16. Denote the sets (8.32) and (8.52) by P and by F , respectively. Since $F \subset P$, $\gcd P \leq \gcd F$. The reverse inequality follows from the fact that each integer in P is a sum of integers in F .
- 8.17. Consider the chain with states $0, 1, \dots$ and α and transition probabilities $p_{0j} = f_{j+1}$ for $j \geq 0$, $p_{0\alpha} = 1 - f$, $p_{i,i-1} = 1$ for $i \geq 1$, and $p_{\alpha\alpha} = 1$ (α is an absorbing state). The transition matrix is

$$\begin{bmatrix} f_1 & f_2 & f_3 & \cdots & 1-f \\ 1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & \cdots & 1 \end{bmatrix}$$

Show that $f_{00}^{(n)} = f_n$ and $p_{00}^{(n)} = u_n$, and apply Theorem 8.1. Then assume $f = 1$, discard the state α and any states j such that $f_k = 0$ for $k > j$, and apply Theorem 8.8.

In FELLER, Volume 1, the renewal theorem is proved by purely analytic means and is then used as the starting point for the theory of Markov chains. Here the procedure is the reverse.

- 8.19. The transition probabilities are $p_{0r} = 1$ and $p_{i,r-i+1} = p$, $p_{i,r-i} = q$, $1 \leq i \leq r$; the stationary probabilities are $u_1 = \cdots = u_r = q^{-1}u_0 = (r+q)^{-1}$. The chance of getting wet is u_0p , of which the maximum is $2r+1-2\sqrt{r(r+1)}$. For $r=5$ this is .046, the pessimal value of p being .523. Of course, $u_0p \leq 1/4r$. In more reasonable climates fewer umbrellas suffice: if $p = .25$ and $r = 3$, then $u_0p = .050$; if $p = .1$ and $r = 2$, then $u_0p = .031$. At the other end of the scale, if $p = .8$ and $r = 3$, then $u_0p = .050$; and if $p = .9$ and $r = 2$, then $u_0p = .043$.
- 8.22. For the last part, consider the chain with state space C_m and transition probabilities p_{ij} for $i, j \in C_m$ (show that they do add to 1).
- 8.23. Let $C' = S - (T \cup C)$, and take $U = T \cup C'$ in (8.51). The probability of absorption in C is the probability of ever entering it, and for initial states i in $T \cup C'$ these probabilities are the minimal solution of

$$y_i = \sum_{j \in T} p_{ij} y_j + \sum_{j \in C} p_{ij} + \sum_{j \in C'} p_{ij} y_j, \quad i \in T \cup C',$$

$$0 \leq y_i \leq 1, \quad i \in T \cup C'.$$

Since the states in C' ($C' = \emptyset$ is possible) are persistent and C is closed, it is impossible to move from C' to C . Therefore, in the minimal solution of the system above, $y_i = 0$ for $i \in C'$. This gives the system (8.55). It also gives, for the minimal solution, $\sum_{j \in T} p_{ij} y_j + \sum_{j \in C} p_{ij} = 0$, $i \in C'$. This makes probabilistic sense: for an i in C' , not only is it impossible to move to a j in C , it is impossible to move to a j in T for which there is positive probability of absorption in C .

- 8.24. Fix on a state i , and let S_ν consist of those j for which $p_{ij}^{(n)} > 0$ for some n congruent to ν modulo t . Choose k so that $p_{ji}^{(k)} > 0$; if $p_{ij}^{(m)}$ and $p_{ji}^{(n)}$ are positive, then t divides $m+k$ and $n+k$, so that m and n are congruent modulo t . The S_ν are thus well defined.
- 8.25. Show that Theorem 8.6 applies to the chain with transition probabilities $p_{ij}^{(t)}$.
- 8.27. (a) From $PC = C\Lambda$ follows $Pc_i = \lambda_i c_i$, from $RP = \Lambda R$ follows $r_i P = \lambda_i r_i$, and from $RC = I$ follows $r_i c_j = \delta_{ij}$. Clearly Λ^n is diagonal and $P^n = C\Lambda^n R$. Hence $p_{ij}^{(n)} = \sum_{uv} C_{iu} \lambda_u^n \delta_{uv} R_{vj} = \sum_u \lambda_u^n (c_u r_u)_{ij} = \sum_u \lambda_u^n (A_u)_{ij}$.
- (b) By Problem 8.26, there are scalars ρ and γ such that $r_1 = \rho r_0 = \rho(\pi_1, \dots, \pi_s)$ and $c_1 = \gamma c_0$, where c_0 is the column vector of 1's. From $r_1 c_1 = 1$ follows $\rho\gamma = 1$, and hence $A_1 = c_1 r_1 = c_0 r_0$ has rows (π_1, \dots, π_s) . Of course, (8.56) gives the exact rate of convergence. It is useful for numerical work; see ÇINLAR, pp. 364 ff.
- (c) Suppose all four p_{ij} are positive. Then $\pi_1 = p_{21}/(p_{21} + p_{12})$, $\pi_2 = p_{12}/(p_{21} + p_{12})$, the second eigenvalue is $\lambda = 1 - p_{12} - p_{21}$, and

$$P^n = \begin{bmatrix} \pi_1 & \pi_2 \\ \pi_1 & \pi_2 \end{bmatrix} + \lambda^n \begin{bmatrix} \pi_2 & -\pi_2 \\ -\pi_1 & \pi_1 \end{bmatrix}.$$

Note that for given n and ϵ , $\lambda^n > 1 - \epsilon$ is possible, which means that the $p_{ij}^{(n)}$ are not yet near the π_j . In the case of positive p_{ij} , P is always diagonalizable by

$$C = \begin{bmatrix} 1 & \pi_2 \\ 1 & -\pi_1 \end{bmatrix}, \quad R = \begin{bmatrix} \pi_1 & \pi_2 \\ 1 & -1 \end{bmatrix}.$$

(d) For example, take $t = \frac{1}{3}$, $0 < \epsilon < t$, and

$$P = \begin{bmatrix} t & t & t \\ t & t & t \\ t - \epsilon & t + \epsilon & t \end{bmatrix}.$$

In this case, 0 is an eigenvalue with algebraic multiplicity 2 and geometric multiplicity 1.

8.30. Show that $\alpha_n = \pi_i$ and

$$\begin{aligned} \beta_n - \alpha_n &= \frac{2}{n(n-1)} \sum_{k=1}^{n-1} \pi_i(n-k)(p_{ii}^{(k)} - \pi_i) \\ &= O\left(\frac{1}{n^2} \sum_{k=1}^n (n-k)\rho^k\right) = O\left(\frac{1}{n}\right), \end{aligned}$$

where ρ is as in Theorem 8.9.

8.36. The definitions give

$$\begin{aligned} E_i[f(X_{\sigma_n})] &= P_i[\sigma_n < n]f(0) + P_i[\sigma_n = n]f(i+n) \\ &= 1 - P_i[\sigma_n = n] + P_i[\sigma_n = n](1 - f_{i+n,0}) \\ &= 1 - p_i \cdots p_{i+n-1} + p_1 \cdots p_{i+n-1}(p_{i+n}p_{i+n+1} \cdots), \end{aligned}$$

and this goes to 1. Since $P_i[\tau \leq n = \sigma_n] \geq P_i[(\tau \leq n) \cap (X_k > 0, k \geq 1)] \rightarrow 1 - f_{i0} > 0$, there is an n of the kind required in the last part of the problem. And now

$$\begin{aligned} E_i[f(X_\tau)] &\leq P_i[\tau < n = \sigma_n]f(i+n) + 1 - P_i[\tau < n = \sigma_n] \\ &= 1 - P_i[\tau < n = \sigma_n]f_{i+n,0}. \end{aligned}$$

8.37. If $i \geq 1$, $n_1 < n_2$, $(i, \dots, i+n_1) \in I_{n_1}$, and $(i, \dots, i+n_2) \in I_{n_2}$, then $P_i[\tau = n_1, \tau = n_2] \geq P_i[X_k = i+k, k \leq n_2] > 0$, which is impossible.

Section 9

9.3. See BAHADUR.

9.7. Because of Theorem 9.6 there are for $P[M_n \geq \alpha]$ bounds of the same order as the ones for $P[S_n \geq \alpha]$ used in the proof of (9.36).

Section 10

- 10.7. Let μ_1 be counting measure on the σ -field of all subsets of a countably infinite Ω , let $\mu_2 = 2\mu_1$, and let \mathcal{P} consist of the cofinite sets. Granted the existence of Lebesgue measure λ on \mathcal{R}^1 , one can construct another example: let $\mu_1 = \lambda$ and $\mu_2 = 2\lambda$, and let \mathcal{P} consist of the half-infinite intervals $(-\infty, x]$.

There are similar examples with a field \mathcal{F}_0 in place of \mathcal{P} . Let Ω consist of the rationals in $(0, 1]$, let μ_1 be counting measure, let $\mu_2 = 2\mu_1$, and let \mathcal{F}_0 consist of finite disjoint unions of “intervals” $[r \in \Omega: a < r \leq b]$.

Section 11

- 11.4. (b) If $(f, g] \subset \bigcup_k (f_k, g_k]$, then $(f(\omega), g(\omega)] \subset \bigcup_k (f_k(\omega), g_k(\omega)]$ for all ω , and Theorem 1.3 gives $g(\omega) - f(\omega) \leq \sum_k (g_k(\omega) - f_k(\omega))$. If $h_m = (g - f - \sum_{k \leq m} (g_k - f_k)) \vee 0$, then $h_n \downarrow 0$ and $g - f \leq \sum_{k \leq n} (g_k - f_k) + h_n$. The positivity and continuity of Λ now give $\nu_0(f, g] \leq \sum_k \nu_0(f_k, g_k]$. A similar, easier argument shows that $\sum_k \nu_0(f_k, g_k] \leq \nu_0(f, g]$ if $(f_k, g_k]$ are disjoint subsets of $(f, g]$.

- 11.5. (b) From (11.7) it follows that $[f > 1] \in \mathcal{F}_0$ for f in \mathcal{L} . Since \mathcal{L} is linear, $[f > x]$ and $[f < -x]$ are in \mathcal{F}_0 for $f \in \mathcal{L}$ and $x > 0$. Since the sets (x, ∞) and $(-\infty, -x)$ for $x > 0$ generate \mathcal{R}^1 , each f in \mathcal{L} is measurable $\sigma(\mathcal{F}_0)$. Hence $\mathcal{F} = \sigma(\mathcal{F}_0)$.

It is easy to show that \mathcal{F}_0 is a semiring and is in fact closed under the formation of proper differences. It can happen that $\Omega \notin \mathcal{F}_0$ —for example, in the case where $\Omega = \{1, 2\}$ and \mathcal{L} consists of the f with $f(1) = 0$. See Jürgen Kindler: A simple proof of the Daniell–Stone representation theorem. *Amer. Math. Monthly*, **90** (1983), 396–397.)

Section 12

- 12.4. (a) If $\theta_n = \theta_m$, then $\theta_{n-m} = 0$ and $n = m$ because θ is irrational. Split G into finitely many intervals of length less than ϵ ; one of them must contain points θ_{2n} and θ_{2m} with $\theta_{2n} < \theta_{2m}$. If $k = m - n$, then $0 < \theta_{2m} - \theta_{2n} = \theta_{2m} \ominus \theta_{2n} = \theta_{2k} < \epsilon$, and the points θ_{2kl} for $1 \leq l \leq \lfloor \theta_{2k}^{-1} \rfloor$ form a chain in which the distance from each to the next is less than ϵ , the first is to the left of ϵ , and the last is to the right of $1 - \epsilon$.

(c) If $s_1 \ominus s_2 = \theta_{2k+1} \ominus \theta_{2n_1} \oplus \theta_{2n_2}$ lies in the subgroup, then $s_1 = s_2$ and $\theta_{2k+1} = \theta_{2(n_1 - n_2)}$.

- 12.5. (a) The $S \oplus \theta_m$ are disjoint, and $(2n + 1)v + k = (2n + 1)v' + k'$ with $|k|, |k'| \leq n$ is impossible if $v \neq v'$.

(b) The $A \oplus \theta_{(2n+1)v'}$ are disjoint, contained in G , and have the same Lebesgue measure.

- 12.6. See Example 2.10 (which applies to any finite measure).

- 12.8. By Theorem 12.3 and Problem 2.19(b), A contains two disjoint compact sets of arbitrarily small positive measure. Construct inductively compact sets $K_{u_1 \dots u_n}$ (each u_i is 0 or 1) such that $0 < \mu(K_{u_1 \dots u_n}) < 3^{-n}$ and $K_{u_1 \dots u_n, 0}$ and $K_{u_1 \dots u_n, 1}$ are disjoint subsets of $K_{u_1 \dots u_n}$. Take $K = \bigcap_n \bigcup_{u_1 \dots u_n} K_{u_1 \dots u_n}$. The Cantor set is a special case.

Section 13

13.3. If $f = \sum_i x_i I_{A_i}$ and $A_i \in T^{-1}\mathcal{F}'$, take A'_i in \mathcal{F}' so that $A_i = T^{-1}A'_i$, and set $\varphi = \sum_i x_i I_{A'_i}$. For the general f measurable $T^{-1}\mathcal{F}'$, there exist simple functions f_n , measurable $T^{-1}\mathcal{F}'$, such that $f_n(\omega) \rightarrow f(\omega)$ for each ω . Choose φ_n , measurable \mathcal{F}' , so that $f_n = \varphi_n T$. Let C' be the set of ω' for which $\varphi_n(\omega')$ has a finite limit, and define $\varphi(\omega') = \lim_n \varphi_n(\omega')$ for $\omega' \in C'$ and $\varphi(\omega') = 0$ for $\omega' \notin C'$. Theorem 20.1(ii) is a special case.

13.7. The class of Borel functions contains the continuous functions and is closed under pointwise passages to the limit and hence contains \mathcal{X} .

By imitating the proof of the π - λ theorem, show that, if f and g lie in \mathcal{X} , then so do $f+g$, fg , $f-g$, $f \vee g$ (note that, for example, $[g: f+g \in \mathcal{X}]$ is closed under passages to the limit). If $f_n(x)$ is 1 or $1 - n(x - \alpha)$ or 0 as $x \leq \alpha$ or $\alpha \leq x \leq \alpha + n^{-1}$ or $\alpha + n^{-1} \leq x$, then f_n is continuous and $f_n(x) \rightarrow I_{(-\infty, \alpha]}(x)$. Show that $[A: I_A \in \mathcal{X}]$ is a λ -system. Conclude that \mathcal{X} contains all indicators of Borel sets, all simple Borel functions, all Borel functions.

13.13. Let $B = \{b_1, \dots, b_k\}$, where $k \leq n$, $E_i = C - b_i^{-1}A$, and $E = \bigcup_{i=1}^k E_i$. Then $E = C - \bigcup_{i=1}^k b_i^{-1}A$. Since μ is invariant under rotations, $\mu(E_i) = 1 - \mu(A) < n^{-1}$, and hence $\mu(E) < 1$. Therefore $C - E = \bigcap_{i=1}^k b_i^{-1}A$ is nonempty. Use any θ in $C - E$.

Section 14

14.3. (b) Since $u \leq F(x)$ is equivalent to $\varphi(u) \leq x$, it follows that $u \leq F(\varphi(u))$. And since $F(x) < u$ is equivalent to $x < \varphi(u)$, it follows further that $F(\varphi(u) - \epsilon) < u$ for positive ϵ .

14.4. (a) If $0 < u < v < 1$, then $P[u \leq F(X) < v, X \in C] = P[\varphi(u) \leq X < \varphi(v), X \in C]$. If $\varphi(u) \in C$, this is at most $P[\varphi(u) \leq X < \varphi(v)] = F(\varphi(v) -) - F(\varphi(u) -) = F(\varphi(v) -) - F(\varphi(u)) \leq v - u$; if $\varphi(u) \notin C$, it is at most $P[\varphi(u) < X < \varphi(v)] = F(\varphi(v) -) - F(\varphi(u)) \leq v - u$. Thus $P[F(X) \in [u, v), X \in C] \leq \lambda[u, v)$ if $0 < u < v < 1$. This is true also for $u = 0$ (let $u \downarrow 0$ and note that $P[F(X) = 0] = 0$) and for $v = 1$ (let $v \uparrow 1$). The finite disjoint unions of intervals $[u, v)$ in $[0, 1)$ form a field there, and by addition $P[F(X) \in A, X \in C] \leq \lambda(A)$ for A in this field. By the monotone class theorem, the inequality holds for all Borel sets in $[0, 1)$. Since $P[F(X) = 1, X \in C] = 0$, this holds also for $A = \{1\}$.

14.5. The sufficiency is easy. To prove necessity, choose continuity points x_i of F in such a way that $x_0 < x_1 < \dots < x_k$, $F(x_0) < \epsilon$, $F(x_k) > 1 - \epsilon$, and $x_i - x_{i-1} < \epsilon$. If n exceeds some n_0 , $|F(x_i) - F_n(x_i)| < \epsilon/2$ for all i . Suppose that $x_{i-1} \leq x \leq x_i$. Then $F_n(x) \leq F_n(x_i) \leq F(x_i) + \epsilon/2 \leq F(x + \epsilon) + \epsilon/2$. Establish a similar inequality going the other direction, and give special arguments for the cases $x \leq x_0$ and $x \geq x_k$.

Section 15

15.1. Suppose there is an \mathcal{F} -partition such that $\sum_i [\sup_{A_i} f] \mu(A_i) < \infty$. Then $a_i = \sup_{A_i} f < \infty$ for i in the set I of indices for which $\mu(A_i) > 0$. If $a = \max_I a_i$, then $\mu[f > a] = \sum_i \mu(A_i \cap [f > a]) \leq \sum_i \mu(A_i \cap [f > a_i]) = 0$. And $A_i \cap [f > 0] = \emptyset$ for i outside the set J of indices for which $\mu(A_i) < \infty$, so that $\mu[f > 0] = \sum \mu(A_i \cap [f > 0]) \leq \sum_J \mu(A_i) < \infty$.

15.4. Let $(\Omega, \mathcal{F}^+, \mu^+)$ be the completion (Problems 3.10 and 10.5) of $(\Omega, \mathcal{F}, \mu)$. If g is measurable \mathcal{F} , $[f \neq g] \subset A$, $A \in \mathcal{F}$, $\mu(A) = 0$, and $H \in \mathcal{R}^1$, then $[f \in H] = (A^c \cap [f \in H]) \cup (A \cap [f \in H]) = (A^c \cap [g \in H]) \cup (A \cap [f \in H])$ lies in \mathcal{F}^+ , and hence f is measurable \mathcal{F}^+ .

(a) Since f is measurable \mathcal{F}^+ , it will be enough to prove that for each (finite) \mathcal{F}^+ -partition $\{B_j\}$ there is an \mathcal{F} -partition $\{A_i\}$ such that $\sum_i [\inf_{A_i} f] \mu(A_i) \geq \sum_j [\inf_{B_j} f] \mu^+(B_j)$, and to prove the dual relation for the upper sums. Choose (Problem 3.2) \mathcal{F} -sets A_i so that $A_i \subset B_i$ and $\mu(A_i) = \mu_*(B_i) = \mu^+(B_i)$. For the partition consisting of the A_i together with $(\cup_i A_i)^c$, the lower sum is at least $\sum_i [\inf_{B_i} f] \mu(A_i) = \sum_i [\inf_{B_i} f] \mu^+(B_i)$.

(b) Choose successively finer \mathcal{F} -partitions $\{A_{ni}\}$ in such a way that the corresponding upper and lower sums differ by at most $1/n^3$. Let g_n and f_n have values $\inf_{A_{ni}} f$ and $\sup_{A_{ni}} f$ on A_{ni} . Use Markov's inequality—since $\mu(\Omega)$ is finite, it may as well be 1—to show that $\mu[f_n - g_n \geq 1/n] \leq 1/n^2$, and then use the first Borel-Cantelli lemma to show that $f_n - g_n \rightarrow 0$ almost everywhere. Take $g = \lim_n g_n$.

Section 16

16.3. $0 \leq f_n - f_1 \uparrow f - f_1$.

16.4. (a) By Fatou's lemma,

$$\begin{aligned} \int f d\mu - \int a d\mu &= \int \lim_n (f_n - a_n) d\mu \\ &\leq \liminf_n \int (f_n - a_n) d\mu = \liminf_n \int f_n d\mu - \int a d\mu \end{aligned}$$

and

$$\begin{aligned} \int b d\mu - \int f d\mu &= \int \lim_n (b_n - f_n) d\mu \\ &\leq \liminf_n \int (b_n - f_n) d\mu = \int b d\mu - \limsup_n \int f_n d\mu. \end{aligned}$$

Therefore

$$\limsup_n \int f_n d\mu \leq \int f d\mu \leq \liminf_n \int f_n d\mu.$$

16.6. For $\omega \in A$ and small enough complex h ,

$$|f(\omega, z_0 + h) - f(\omega, z_0)| = \left| \int_{z_0}^{z_0+h} f'(\omega, z) dz \right| \leq |h| g(\omega, z_0).$$

16.8. Use the fact that $\int_A |f| d\mu \leq \alpha \mu(A) + \int_{\{|f| \geq \alpha\}} |f| d\mu$.

- 16.9.** If $\mu(A) < \delta$ implies $\int_A |f_n| d\mu < \epsilon$ for all n , and if $\alpha^{-1} \sup_n \int |f_n| d\mu < \delta$, then $\mu[|f_n| \geq \alpha] \leq \alpha^{-1} \int |f_n| d\mu < \delta$ and hence $\int_{[|f_n| \geq \alpha]} |f_n| d\mu < \epsilon$ for all n . For the reverse implication adapt the argument in the preceding note.
- 16.10.** (b) Suppose that f_n are nonnegative and satisfy condition (ii) and μ is nonatomic. Choose δ so that $\mu(A) \leq \delta$ implies $\int_A f_n d\mu \leq 1$ for all n . If $\mu[f_n = \infty] > 0$, there is an A such that $A \subset [f_n = \infty]$ and $0 < \mu(A) < \delta$; but then $\int_A f_n d\mu = \infty$. Since $\mu[f_n = \infty] = 0$, there is an α such that $\mu[f_n > \alpha] \leq \delta \leq \mu[f_n \geq \alpha]$. Choose $B \subset [f_n = \alpha]$ in such a way that $A = [f_n > \alpha] \cup B$ satisfies $\mu(A) = \delta$. Then $\alpha\delta = \alpha\mu(A) \leq \int_A f_n d\mu \leq 1$ and $\int f_n d\mu \leq 1 + \alpha\mu(A^c) \leq 1 + \delta^{-1}\mu(\Omega)$.
- 16.12.** (b) Suppose that $f \in \mathcal{L}$ and $f \geq 0$. If $f_n = (1 - n^{-1})f \vee 0$, then $f_n \in \mathcal{L}$ and $f_n \uparrow f$, so that $\nu(f_n, f) = \Lambda(f - f_n) \downarrow 0$. Since $\nu(f_1, f) < \infty$, it follows that $\nu[(\omega, t): f(\omega) = t] = 0$. The disjoint union

$$B_n = \bigcup_{i=1}^{n2^n} \left(\left[\frac{i}{2^n} < f \leq \frac{i+1}{2^n} \right] \times \left(0, \frac{i}{2^n} \right] \right)$$

increases to B , where $B \subset (0, f]$ and $(0, f] - B \subset [(\omega, t): f(\omega) = t]$. Therefore

$$\Lambda(f) = \nu(0, f] = \lim_n \nu(B_n) = \lim_n \sum_{i=1}^{n2^n} \frac{i}{2^n} \mu \left[\frac{i}{2^n} < f \leq \frac{i+1}{2^n} \right] = \int f d\mu.$$

Section 17

- 17.1.** (a) Let A_ϵ be the set of x such that for every δ there are points y and z satisfying $|y - x| < \delta$, $|z - x| < \delta$, and $|f(y) - f(z)| \geq \epsilon$. Show that A_ϵ is closed and D_f is the union of the A_ϵ .
- (c) Given ϵ and η , choose a partition into intervals I_i for which the corresponding upper and lower sums differ by at most $\epsilon\eta$. By considering those I_i whose interiors meet A_ϵ , show that $\epsilon\eta \geq \epsilon\lambda(A_\epsilon)$.
- (d) Let M bound $|f|$ and, given ϵ , find an open G such that $D_f \subset G$ and $\lambda(G) < \epsilon/M$. Take $C = [0, 1] - G$ and show by compactness that there is a δ such that $|f(y) - f(x)| < \epsilon$ if x (but perhaps not y) lies in C and $|y - x| < \delta$. If $[0, 1]$ is decomposed into intervals I_i with $\lambda(I_i) < \delta$, and if $x_i \in I_i$, let g be the function with value $f(x_i)$ on I_i . Let Σ' denote summation over those i for which I_i meets C , and let Σ'' denote summation over the other i . Show that

$$\begin{aligned} \left| \int_0^1 f(x) dx - \sum f(x_i) \lambda(I_i) \right| &\leq \int_0^1 |f(x) - g(x)| dx \\ &\leq \Sigma' 2\epsilon \lambda(I_i) + \Sigma'' 2M \lambda(I_i) < 4\epsilon. \end{aligned}$$

- 17.10.** (c) Do not overlook the possibility that points in $(0, 1) - K$ converge to a point in K .
- 17.11.** (b) Apply the bounded convergence theorem to $f_n(x) = (1 - n \operatorname{dist}(x, [s, t]))^+$.
- (c) The class of Borel sets B in $[u, v]$ for which $f = I_B$ satisfies (17.8) is a λ -system.

(e) Choose simple f_n such that $0 \leq f_n \uparrow f$. To (17.8) for $f = f_n$, apply the monotone convergence theorem on the right and the dominated convergence theorem on the left.

- 17.12. If $g(x)$ is the distance from x to $[a, b]$, then $f_n = (1 - ng) \vee 0 \downarrow I_{[a, b]}$ and $f_n \in \mathcal{L}$; since the continuous functions are measurable \mathcal{R}^1 , it follows that $\mathcal{F} = \mathcal{R}^1$. If $f_n(x) \downarrow 0$ for each x , then the compact sets $[x: f_n(x) \geq \epsilon]$ decrease to \emptyset and hence one of them is \emptyset ; thus the convergence is uniform.
- 17.13. The linearity and positivity of Λ are certainly elementary facts, and for the continuity property, note that if $0 \leq f \leq \epsilon$ and f vanishes outside $[a, b]$, then elementary considerations show that $0 \leq \Lambda(f) \leq \epsilon(b - a)$.

Section 18

- 18.2. First, $\mathcal{A} \times \mathcal{A}$ is generated by the sets of the forms $\{x\} \times X$ and $X \times \{x\}$. If the diagonal E lies in $\mathcal{A} \times \mathcal{A}$, then there must be a countable S in X such that E lies in the σ -field \mathcal{F} generated by the sets of these two forms for x in S . If \mathcal{P} consists of S^c and the singletons in S , then \mathcal{F} is the class of unions of sets in the partition $[P_1 \times P_2: P_1, P_2 \in \mathcal{P}]$. But $E \in \mathcal{F}$ is impossible.
- 18.3. Consider $A \times B$, where A consists of a single point and B lies outside the completion of \mathcal{R}^1 with respect to λ .
- 18.17. Put $f_p = p^{-1} \log p$, and put $f_n = 0$ if n is not a prime. In the notation of (18.17), $F(x) = \log x + \varphi(x)$, where φ is bounded because of (5.51). If $G(x) = -1/\log x$, then

$$\begin{aligned} \sum_{p \leq x} \frac{1}{p} &= \frac{F(x)}{\log x} + \int_2^x \frac{F(t) dt}{t \log^2 t} \\ &= 1 + \frac{\varphi(x)}{\log x} + \int_2^x \frac{dt}{t \log t} + \int_2^\infty \frac{\varphi(t) dt}{t \log^2 t} - \int_x^\infty \frac{\varphi(t) dt}{t \log^2 t}. \end{aligned}$$

Section 19

- 19.3. See BANACH, p. 34.
- 19.4. (a) Take $f = 0$ and $f_n = I_{(0, 1/n)}$.
 (b) Take $f = 0$, and let $\{f_n\}$ be an infinite orthonormal set. Use the fact that $\sum_n (f_n, g)^2 \leq \|g\|^2$.
- 19.5. Take $f_n = nI_{(0, 1/n)}$, and suppose that f_{n_k} converges weakly to some f in L^1 . Integrate against the L^∞ -functions $\operatorname{sgn} f \cdot I_{(\epsilon, 1)}$ and conclude that $f = 0$ almost everywhere; now integrate against the function identically 1 and get a contradiction.

Section 20

- 20.4. Suppose U_1, \dots, U_k are independent and uniformly distributed over the unit interval, put $V_i = 2nU_i - n$, and let μ_n be $(2n)^k$ times the distribution of

(V_1, \dots, V_k) . Then μ_n is supported by $Q_n = (-n, n] \times \cdots \times (-n, n]$, and if $I = (a_1, b_1] \times \cdots \times (a_k, b_k] \subset Q_n$, then $\mu_n(I) = \prod_{i=1}^k (b_i - a_i)$. Further, if $A \subset Q_n \subset Q_m$ ($n < m$), then $\mu_n(A) = \mu_m(A)$. Define $\lambda_k(A) = \lim_n \mu_n(A \cap Q_n)$.

20.7. By the argument preceding (8.16),

$$P_i[T_1 = n_1, \dots, T_k = n_k] = f_{ij}^{(n_1)} f_{jj}^{(n_2 - n_1)} \cdots f_{jj}^{(n_k - n_{k-1})}.$$

For the general initial distribution, average over i .

- 20.8. (a) Use the π - λ theorem to show that $P[(X_{\pi 1}, \dots, X_{\pi n}) \in H]$ is the same for all permutations π .
(b) Use part (a) and the fact that $Y_n = r$ if and only if $T_r^{(n)} = n$.
(c) If $k \leq n$, then $Y_k = r$ if and only if exactly $r - 1$ among the integers $1, \dots, k - 1$ precede k in the permutation $T^{(n)}$.
(d) Observe that $T^{(n)} = (t_1, \dots, t_n)$ and $Y_{n+1} = r$ if and only if $T^{(n+1)} = (t_1, \dots, t_{r-1}, n + 1, t_r, \dots, t_n)$, and conclude that $\sigma(Y_{n+1})$ is independent of $\sigma(T^{(n)})$ and hence of $\sigma(Y_1, \dots, Y_n)$ —see Problem 20.6.

20.12. If X and Y are independent, then

$$P[|(X + Y) - (x + y)| < \epsilon] \geq P[|X - x| < \tfrac{1}{2}\epsilon] P[|Y - y| < \tfrac{1}{2}\epsilon]$$

and

$$P[X + Y = x + y] \geq P[X = x] P[Y = y].$$

20.14. The partial-fraction expansion gives

$$c_u(y - x)c_v(x) = \frac{uv}{\pi^2} \frac{1}{R} (A + B + C + D),$$

where $R = (u^2 - v^2)^2 + 2(u^2 + v^2)y^2 + y^4$ and

$$\begin{aligned} A &= \frac{y^2 + v^2 - u^2}{u^2 + (y - x)^2}, & B &= \frac{2y(y - x)}{u^2 + (y - x)^2}, \\ C &= \frac{y^2 - v^2 + u^2}{v^2 + x^2}, & D &= \frac{2yx}{v^2 + x^2}. \end{aligned}$$

After the fact this can of course be checked mechanically. Integrate over $[-t, t]$ and let $t \rightarrow \infty$: $\int_{-t}^t D \, dx = 0$, $\int_{-t}^t B \, dx \rightarrow 0$, and $\int_{-\infty}^{\infty} (A + C) \, dx = (y^2 + v^2 - u^2)u^{-1}\pi + (y^2 - v^2 + u^2)v^{-1}\pi = u^{-1}v^{-1}\pi^2 R c_{u+v}(y)$. There is a very simple proof by characteristic functions; see Problem 26.9.

- 20.16. See Example 20.1 for the case $n = 1$, prove by inductive convolution and a change of variable that the density must have the form $K_n x^{(n/2)-1} e^{-x/2}$, and then from the fact that the density must integrate to 1 deduce the form of K_n .
20.17. Show by (20.38) and a change of variable that the left side of (20.48) is some constant times the right side; then show that the constant must be 1.

- 20.20.** (a) Given ϵ choose M so that $P[|X| > M] < \epsilon$ and $P[|Y| > M] < \epsilon$, and then choose δ so that $|x|, |y| \leq M$, $|x - x'| < \delta$, and $|y - y'| < \delta$ imply that $|f(x', y') - f(x, y)| < \epsilon$. Note that $P[|f(X_n, Y_n) - f(X, Y)| \geq \epsilon] \leq 2\epsilon + P[|X_n - X| \geq \delta] + P[|Y_n - Y| \geq \delta]$.
- 20.23.** Take, for example, independent X_n assuming the values 0 and n with probabilities $1 - n^{-1}$ and n^{-1} . Estimate the probability that $X_k = k$ for some k in the range $n/2 < k \leq n$.
- 20.24.** (b) For each m split A into 2^m sets A_{mk} of probability $P(A)/2^m$. Arrange all the A_{mk} in one infinite sequence, and let X_n be the indicator of the n th set in it.
- 20.27.** To get the distribution of Φ , show by integration that for $0 \leq \phi \leq 2\pi$, the intersection with the unit ball of the (x_1, x_2, x_3) -set where $0 \leq x_3 \leq (x_1^2 + x_2^2)^{1/2} \tan \phi$ has volume $\frac{2}{3}\pi \sin \phi$.

Section 21

- 21.5.** Consider $\sum I_{A_n}$. A random variable is finite with probability 1 if (but not only if) it is integrable.
- 21.6.** Calculate $\int_0^\infty x dF(x) = \int_0^\infty \int_0^x dy dF(x) = \int_0^\infty \int_y^\infty dF(x) dy$.
- 21.8.** (a) Write $E[Y - X] = \int_{X < Y} \int_{X < t \leq Y} dt dP - \int_{Y < X} \int_{Y < t \leq X} dt dP$.
- 21.10.** (a) The most important dependent uncorrelated random variables are the trigonometric functions—the random variables $\sin 2\pi n\omega$ and $\cos 2\pi n\omega$ on the unit interval with Lebesgue measure. See Problem 19.8.
- 21.13.** Use Fubini's theorem; see (20.29) and (20.30).
- 21.14.** Even if $X = -Y$ is not integrable, $X + Y = 0$ is. Since $|Y| \leq |x| + |x + Y|$, $E[|Y|] = \infty$ implies that $E[|x + Y|] = \infty$ for each x ; use Problem 21.13. See also the lemma in Section 28.

21.21. Use (21.12).

Section 22

- 22.2.** For sufficiency, use $E[\sum |X_n^{(c)}|] = \sum E[|X_n^{(c)}|]$.
- 22.8.** (a) Put $U = \sum_k I_{[k \leq \tau]} X_k^+$ and $V = \sum_k I_{[k \leq \tau]} X_k^-$, so that $S_\tau = U - V$. Since $[\tau \geq k] = \Omega - [\tau \leq k-1]$ lies in $\sigma(X_1, \dots, X_{k-1})$, it follows that $E[I_{[\tau \geq k]} X_k^+] = E[I_{[\tau \geq k]}] E[X_k^+] = P[\tau \geq k] E[X_1^+]$. Hence $E[U] = \sum_{k=1}^\infty E[X_1^+] P[\tau \geq k] = E[X_1^+] E[\tau]$. Treat V the same way.
- (b) To prove $E[\tau] < \infty$, show that $P[\tau > (a+b)n] \leq (1 - \rho^{a+b})^n$. By (7.7), S_τ is b with probability $(1 - \rho^a)/(1 - \rho^{a+b})$ and $-a$ with the opposite probability. Since $E[X_1] = p - q$,

$$E[\tau] = \frac{a}{q-p} - \frac{a+b}{q-p} \frac{1-\rho^a}{1-\rho^{a+b}}, \quad \rho = \frac{q}{p} \neq 1.$$

22.11. For each θ , $\sum_n e^{iX_n}(e^{i\theta}z)^n$ has the same probabilistic behavior as the original series, because the $X_n + n\theta$ reduced modulo 2π are independent and uniformly distributed. Therefore, the rotation idea in the proof of Theorem 22.9 carries over. See KAHANE for further results.

22.14. (b) Let $A = f^{-1}B$ and suppose p is a period of f . Let $m = \lfloor x/p \rfloor$ and $n = \lfloor 1/p \rfloor$. By periodicity, $P(A \cap [y, y+p])$ is the same for all y ; therefore, $|P(A \cap [0, x]) - mP(A \cap [0, p])| \leq p$, $|P(A) - nP(A \cap [0, p])| \leq p$, and $|P(A \cap [0, x]) - P(A)x| \leq 2p + |x - m/n| \leq 3p$. Since p can be taken arbitrarily small, $P(A \cap [0, x]) = P(A)x$.

22.15. (a) By the inequalities $L(s) \leq M(s)$ and (22.24),

$$B_E(2s) = 1 \wedge 3L(2s) \leq 3M(2s) \leq 3B_O(s).$$

For the other inequality, note first that $T(s)$ is nonincreasing and that $R(2s) \leq 2L(s)$ and $T(s) \leq L(s) \leq M(s)$. If $B_E(s) \leq 1/3$, then

$$\begin{aligned} B_O(6s) &\leq \frac{T(6s)}{1 - R(6s)} \leq \frac{T(3s)}{1 - 2L(3s)} \leq \frac{M(3s)}{1 - 2M(3s)} \\ &\leq \frac{B_E(s)}{1 - 2B_E(s)} \leq 3B_E(s). \end{aligned}$$

On the other hand, if $B_E(s) > 1/3$, then $B_O(6s) \leq 1 < 3B_E(s)$. In either case, $B_O(6s) \leq 3B_E(s)$.

Section 23

23.3. Note that A_t cannot exceed t . If $0 \leq u \leq t$ and $v \geq 0$, then $P[A_t \geq u, B_t > v] = P[N_{t+v} - N_{t-u} = 0] = e^{-\alpha u} e^{-\alpha v}$.

23.4. (a) Use (20.37) and the distributions of A_t and B_t .

(b) A long interarrival interval has a better chance of covering t than a short one does.

23.6. The probability that $N'_{S_n+k} - N'_{S_n} = j$ is

$$\int_0^\infty e^{-\beta x} \frac{(\beta x)^j}{j!} \frac{\alpha^k}{\Gamma(k)} x^{k-1} e^{-\alpha x} dx = \frac{\alpha^k \beta^j}{(\alpha + \beta)^{k+j}} \frac{(j+k-1)!}{j!(k-1)!}.$$

23.8. Let M_t be the given process and put $\varphi(t) = E[M_t]$. Since there are no fixed discontinuities, $\varphi(t)$ is continuous. Let $\psi(u) = \inf\{t: u \leq \varphi(t)\}$, and show that $N_u = M_{\psi(u)}$ is an ordinary Poisson process and $M_t = N_{\varphi(t)}$.

23.9. Let $t \rightarrow \infty$ in

$$\frac{S_{N_t}}{N_t} \leq \frac{t}{N_t} \leq \frac{S_{N_t+1}}{N_t+1} \frac{N_t+1}{N_t}.$$

- 23.11. Restrict t in Problem 23.10 to integers. The waiting times are the Z_n of Problem 20.7, and account must be taken of the fact that the distribution of Z_1 may differ from that of the other Z_n .

Section 25

- 25.1. (e) Let G be an open set that contains the rationals and satisfies $\lambda(G) < \frac{1}{2}$. For $k = 0, 1, \dots, n-1$, construct a triangle whose base contains k/n and is contained in G : make these bases so narrow that they do not overlap, and adjust the heights of the triangles so that each has area $1/n$. For the n th density, piece together these triangular functions, and for the limit density, use the function identically 1 over the unit interval.
- 25.2. By Problem 14.8 it suffices to prove that $F_n(\cdot, \omega) \Rightarrow F$ with probability 1, and for this it is enough that $F_n(x, \omega) \rightarrow F(x)$ with probability 1 for each rational x .
- 25.3. (b) It can be shown, for example, that (25.14) holds for $x_n = n!$. See Persi Diaconis: The distribution of leading digits and uniform distribution mod 1, *Ann. Prob.*, 5 (1977), 72–81.
- (c) The first significant digits of numbers drawn at random from empirical compilations such as almanacs and engineering handbooks seem approximately to follow the limiting distribution in (25.15) rather than the uniform distribution over $1, 2, \dots, 9$. This is sometimes called *Benford's law*. One explanation is that the distribution of the observation X and hence of $\log_{10} X$ will be spread over a large interval; if $\log_{10} X$ has a reasonably smooth density, it then seems plausible that $\{\log_{10} X\}$ should be approximately uniformly distributed. See FELLER, Volume 2, p. 62.
- 25.9. Use Scheffé's theorem.
- 25.10. Put $f_n(x) = P[X_n = \gamma_n + k\delta_n]\delta_n^{-1}$ for $\gamma_n + k\delta_n < x \leq \gamma_n + (k+1)\delta_n$. Construct random variables Y_n with densities f_n , and first prove $Y_n \Rightarrow X$. Show that $Z_n = \gamma_n + [(Y_n - \gamma_n)/\delta_n]\delta_n$ has the distribution of X_n and that $Y_n - Z_n \Rightarrow 0$.
- 25.11. For a proof of (25.16) see FELLER, Volume 1, Chapter 7.
- 25.13. (b) Follow the proof of Theorem 25.8, but approximate $I_{(x, y]}$ instead of $I_{(-\infty, x]}$.
- 25.20. Let X_n assume the values n and 0 with probabilities $p_n = 1/(n \log n)$ and $1 - p_n$.

Section 26

- 26.1. (b) Let μ be the distribution of X . If $|\varphi(t)| = 1$ and $t \neq 0$, then $\varphi(t) = e^{ita}$ for some a , and $0 = \int_{-\infty}^{\infty} (1 - e^{it(x-a)})\mu(dx) = \int_{-\infty}^{\infty} (1 - \cos t(x-a))\mu(dx)$. Since the integral vanishes, μ must confine its mass to the points where the nonnegative integrand vanishes, namely to the points x for which $t(x-a) = 2\pi n$ for some integer n .
- (c) The mass of μ concentrates at points of the form $a + 2\pi n/t$ and also at points of the form $a' + 2\pi n/t'$. If μ is positive at two distinct points, it follows that t/t' is rational.

26.3. (a) Let $f_0(x) = \pi^{-1}x^{-2}(1 - \cos x)$ be the density corresponding to $\varphi_0(t)$. If $p_k = (s_k - s_{k+1})t_k$, then $\sum_{k=1}^{\infty} p_k = 1$; since $\sum_{k=1}^{\infty} p_k \varphi_0(t/t_k) = \varphi(t)$ (check the points $t = t_j$), $\varphi(t)$ is the characteristic function of the continuous density $\sum_{k=1}^{\infty} p_k t_k f_0(t_k, x)$.

(b) If $\lim_{t \rightarrow \infty} \varphi(t) = 0$, approximate φ by functions of the kind in part (a), pass to the limit, and use the first corollary to the continuity theorem. If φ does not vanish at infinity, mix in a unit mass at 0.

26.12. On the right in (26.30) replace $\varphi(t)$ by the integral defining it and apply Fubini's theorem; the integral average comes to

$$\mu\{a\} + \int_{x \neq a} \frac{\sin T(x-a)}{T(x-a)} \mu(dx).$$

Now use the bounded convergence theorem.

26.15. (a) Use (26.4₀) to prove that $|\varphi_n(t+h) - \varphi_n(t)| \leq 2\mu_n(-a, a)^c + a|h|$.

(b) Use part (a).

26.17. (a) Use the second corollary to the continuity theorem.

26.19. For the Weierstrass approximation theorem, see RUDIN₁, Theorem 7.32.

26.22. (a) If a_n goes to 0 along a subsequence, then $|\psi(t)| \equiv 1$; use part (c) of Problem 26.1.

(c) Suppose two subsequences of $\{a_n\}$ converge to a_0 and a , where $0 < a_0 < a$; put $\theta = a_0/a$ and show that $|\varphi(t)| = |\varphi(\theta^k t)|$.

(d) Observe that

$$b_n = -i[e^{itb_n} - 1] \left[\int_0^t e^{isb_n} ds \right]^{-1}.$$

26.25. First do the nonnegative case; then note that if f and g have the same coefficients, so do $f^+ + g^-$ and $g^+ + f^-$.

Section 27

27.8. By the same reasoning as in Example 27.3, $(R_n - \log n)/\sqrt{\log n} \Rightarrow N$.

27.9. The Lindeberg theorem applies: $(S_n - n^2/4)/\sqrt{n^3/36} \Rightarrow N$.

27.11. Let Y_n be X_n or 0 according as $|X_n| \leq n^{1/2} \log n$ or not. Show that $X_n = Y_n$ for large n , with probability 1, and that Lyapounov's theorem ($\delta = 1$) applies to the Y_n .

27.12. For example, let the distribution of X_n be the mixture, with weights $1 - n^{-2}$ and n^{-2} , of the standard normal and Cauchy distributions.

27.16. Write $\int_x^\infty e^{-u^2/2} du = x^{-1}e^{-x^2/2} - \int_x^\infty u^{-2}e^{-u^2/2} du$.

27.17. For another approach to large-deviation theory, see Mark Pinsky: An elementary derivation of Khintchine's estimate for large deviations, *Proc. Amer. Math. Soc.*, **22** (1969), 288–290.

27.19. (a) Everything comes from (4.7). If $A = [(l_1, \dots, l_k) \in H]$ and $B \in \sigma(l_{k+n}, l_{k+n+1}, \dots)$, then

$$|P(A \cap B) - P(A)P(B)| \\ \leq \sum |P([l_u = i_u, u \leq k] \cap B) - P[l_u = i_u, u \leq k]P(B)|,$$

where the sum extends over the k -tuples (i_1, \dots, i_k) of nonnegative integers in H . The summand vanishes if $u + i_u < k + n$ for $u \leq k$; the remaining terms add to at most $2\sum_{u=1}^k P[l_u \geq k + n - u] \leq 4/2^n$.

(b) To show that $\sigma^2 = 6$ (see (27.20)), show that l_1 has mean 1 and variance 2 and that

$$\int_{[l_1=i]} l_1 l_{1+n} dP = \begin{cases} P[l_1=i]iE[l_{1+n}] & \text{if } i < n, \\ P[l_1=i]i(i-n) & \text{if } i \geq n. \end{cases}$$

Section 28

28.2. (b) Pass to a subsequence along which $\mu_n(R^1) \rightarrow \infty$, choose ϵ_n so that it decreases to 0 and $\epsilon_n \mu_n(R^1) \rightarrow \infty$, and choose x_n so that it increases to ∞ and $\mu_n(-x_n, x_n) > \frac{1}{2}\mu_n(R^1)$; consider the f that satisfies $f(\pm x_n) = \epsilon_n$ for all n and is defined by linear interpolation in between these points.

28.4. (a) If all functions (28.12) are characteristic functions, they are all certainly infinitely divisible. Since (28.12) is continuous at 0, it need only be exhibited as a limit of characteristic functions. If μ_n has density $I_{[-n,n]}(1+x^2)$ with respect to ν , then

$$\exp \left[i\gamma t + it \int_{-\infty}^{\infty} \frac{x}{1+x^2} \mu_n(dx) + \int_{-\infty}^{\infty} (e^{itx} - 1 - itx) \frac{1}{x^2} \mu_n(dx) \right]$$

is a characteristic function and converges to (28.12). It can also be shown that every infinitely divisible distribution (no moments required) has characteristic function of the form (28.12); see GNEDENKO & KOLMOGOROV, p. 76.

(b) Use (see Problem 18.19) $-|t| = \pi^{-1} \int_{-\infty}^{\infty} (\cos tx - 1)x^{-2} dx$.

28.14. If X_1, X_2, \dots are independent and have distribution function F , then $(X_1 + \dots + X_n)/\sqrt{n}$ also has distribution function F . Apply the central limit theorem.

28.15. The characteristic function of Z_n is

$$\exp \frac{c}{n} \sum_k \frac{1}{(|k|/n)^{1+\alpha}} (e^{itk/n} - 1) \rightarrow \exp c \int_{-\infty}^{\infty} \frac{e^{itx} - 1}{|x|^{1+\alpha}} dx \\ = \exp \left[-c|t|^\alpha \int_{-\infty}^{\infty} \frac{1 - \cos x}{|x|^{1+\alpha}} dx \right].$$

Section 29

- 29.1.** (a) If f is lower semicontinuous, $[x: f(x) > t]$ is open. If f is positive, which is no restriction, then $\int f d\mu = \int_0^\infty \mu[f > t] dt \leq \int_0^\infty \liminf_n \mu_n[f > t] dt \leq \liminf_n \int_0^\infty \mu_n[f > t] dt = \liminf_n \int f d\mu_n$.
- (b) If G is open, then I_G is lower semicontinuous.

- 29.7.** Let Σ be the covariance matrix. Let M be an orthogonal matrix such that the entries of $M\Sigma M'$ are 0 except for the first r diagonal entries, which are 1. If $Y = MX$, then Y has covariance matrix $M\Sigma M'$, and so $Y = (Y_1, \dots, Y_r, 0, \dots, 0)$, where Y_1, \dots, Y_r are independent and have the standard normal distribution. But $|X|^2 = \sum_{i=1}^k Y_i^2$.

- 29.8.** By Theorem 29.5, X_n has asymptotically the centered normal distribution with covariances σ_{ij} . Put $x = (p_1^{1/2}, \dots, p_k^{1/2})$ and show that $\Sigma x' = 0$, so that 0 is an eigenvalue of Σ . Show that $\Sigma y' = y'$ if y is perpendicular to x , so that Σ has 1 as an eigenvalue of multiplicity $k - 1$. Use Problem 29.7 together with Theorem 29.2 ($h(x) = |x|^2$).

- 29.9.** (a) Note that $n^{-1} \sum_{i=1}^n Y_{ni}^2 \Rightarrow 1$ and that (X_{n1}, \dots, X_{nt}) has the same distribution as $(Y_{n1}, \dots, Y_{nt}) / (n^{-1} \sum_{i=1}^n Y_{ni}^2)^{1/2}$.

Section 30

- 30.1.** Rescale so that $s_n^2 = 1$, and put $L_n(\epsilon) = \sum_k \int_{|X_{nk}| \geq \epsilon} X_{nk}^2 dP$. Choose increasing n_u so that $L_n(u^{-1}) \leq u^{-3}$ for $n \geq n_u$, and put $M_n = u^{-1}$ for $n_u \leq n < n_{u+1}$. Then $M_n \rightarrow 0$ and $L_n(M_n) \leq M_n^3$. Put $Y_{nk} = X_{nk} I_{[|X_{nk}| \leq M_n]}$. Show that $\sum_k E[Y_{nk}] \rightarrow 0$ and $\sum_k E[Y_{nk}^2] \rightarrow 1$, and apply to $\sum_k Y_{nk}$ the central limit theorem under (30.5). Show that $\sum_k P[X_{nk} \neq Y_{nk}] \rightarrow 0$.

- 30.4.** Suppose that the moment generating function M_n of μ_n converges to the moment generating function M of μ in some interval about s . Let ν_n have density $e^{sx}/M_n(s)$ with respect to μ_n , and let ν have density $e^{sx}/M(s)$ with respect to μ . Then the moment generating function of ν_n converges to that of ν in some interval about 0, and hence $\nu_n \Rightarrow \nu$. Show that $\int_{-\infty}^\infty f(x) \mu_n(dx) \rightarrow \int_{-\infty}^\infty f(x) \mu(dx)$ if f is continuous and has bounded support; see Problem 25.13(b).

- 30.5.** (a) By Hölder's inequality $|\sum_{j=1}^k t_j x_j|^r \leq k^{r-1} \sum_{j=1}^k |t_j x_j|^r$, and so $\sum_r \theta^r / |\sum_j t_j x_j|^r \mu(dx) / r!$ has positive radius of convergence. Now

$$\int_{R^k} \left(\sum_{j=1}^k t_j x_j \right)^r \mu(dx) = \sum t_1^{r_1} \cdots t_k^{r_k} \alpha(r_1, \dots, r_k),$$

where the summation extends over k -tuples that add to r . Project μ to the line by the mapping $\sum_j t_j x_j$, apply Theorem 30.1, and use the fact that μ is determined by its values on half-spaces.

- 30.6.** Use the Cramér–Wold idea.

30.8. Suppose that $k = 2$ in (30.30). Then

$$\begin{aligned} & M[(\cos \lambda_1 x)^{r_1} (\cos \lambda_2 x)^{r_2}] \\ &= M\left[\left(\frac{e^{i\lambda_1 x} + e^{-i\lambda_1 x}}{2}\right)^{r_1} \left(\frac{e^{i\lambda_2 x} + e^{-i\lambda_2 x}}{2}\right)^{r_2}\right] \\ &= 2^{-r_1-r_2} \sum_{j_1=0}^{r_1} \sum_{j_2=0}^{r_2} \binom{r_1}{j_1} \binom{r_2}{j_2} M[\exp i(\lambda_1(2j_1 - r_1) + \lambda_2(2j_2 - r_2))x]. \end{aligned}$$

By (26.33) and the independence of λ_1 and λ_2 , the last mean here is 1 if $2j_1 - r_1 = 2j_2 - r_2 = 0$ and is 0 otherwise. A similar calculation for $k = 1$ gives (30.28), and a similar calculation for general k gives (30.30). The actual form of the distribution in (30.29) is unimportant. For (30.31) use the multidimensional method of moments (Problem 30.6) and the mapping theorem. For (30.32) use the central limit theorem; by (30.28), X_1 has mean 0 and variance $\frac{1}{2}$.

30.10. If $n^{1/2} < m \leq n$ and the inequality in (30.33) holds, then $\log \log n^{1/2} < \log \log n - \epsilon(\log \log n)^{1/2}$, which implies $\log \log n < \epsilon^{-2} \log^2 2$. For large n the probability in (30.33) is thus at most $1/\sqrt{n}$.

Section 31

31.1. Consider the argument in Example 31.1. Suppose that F has a nonzero derivative at x , and let I_n be the set of numbers whose base- r expansions agree in the first n places with that of x . The analogue of (31.16) is $P[X \in I_{n+1}]/P[X \in I_n] \rightarrow r^{-1}$, and the ratio here is one of p_0, \dots, p_r . If $p_i \neq r^{-1}$ for some i , use the second Borel–Cantelli lemma to show that the ratio is p_i infinitely often except on a set of Lebesgue measure 0. (This last part of the argument is unnecessary if $r = 2$.)

The argument in Example 31.3 needs no essential change. The analogue of (31.17) is

$$F(x) = p_0 + \dots + p_{i-1} + p_i F(rx - i), \quad \frac{i}{r} \leq x \leq \frac{i+1}{r}, \quad 0 \leq i < r-1.$$

31.3. (b) Take $f_1 = I_{g^{-1}H_0}$ and $f_2 = F$; $(f_1 f_2)^{-1}\{1\} = H_0$ is not a Lebesgue set.

31.9. Suppose that A is bounded, define μ by $\mu(B) = \lambda(B \cap A)$, and let F be the corresponding distribution function. It suffices to show that $F'(x) = 1$ for x in A , apart from a set of Lebesgue measure 0. Let C_ϵ be the set of x in A for which $F'(x) \leq 1 - \epsilon$. From Theorem 31.4(i) deduce that $\lambda(C_\epsilon) = \mu(C_\epsilon) \leq (1 - \epsilon)\lambda(C_\epsilon)$ and hence $\lambda(C_\epsilon) = 0$. Thus $F'(x) > 1 - \epsilon$ almost everywhere on A . Obviously, $F'(x) \leq 1$.

31.11. Let A be the set of x in the unit interval for which $F'(x) = 0$, take $\alpha = 0$, and define A_n as in the first part of the proof of Theorem 31.4. Choose n so that $\lambda(A_n) \geq 1 - \epsilon$. Split $\{1, 2, \dots, n\}$ into the set M of k for which $((k-1)/n, k/n]$ meets A_n and the opposite set N . Prove successively that $\sum_{k \in M} [F(k/n) - F((k-1)/n)] \leq \epsilon$, $\sum_{k \in N} [F(k/n) - F((k-1)/n)] \geq 1 - \epsilon$, $\sum_{k \in M} 1/n \geq \lambda(A_n) \geq 1 - \epsilon$, $\sum_{k=1}^n |f(k/n) - f((k-1)/n)| \geq 2 - 2\epsilon$.

31.15. $\prod_{n=1}^{\infty} (\frac{1}{2} + \frac{1}{2}e^{2it/3^n}).$

31.18. For x fixed, let u_n and v_n be the pair of successive dyadic rationals of order n ($v_n - u_n = 2^{-n}$) for which $u_n < x \leq v_n$. Show that

$$\frac{f(v_n) - f(u_n)}{v_n - u_n} = \sum_{k=0}^{n-1} \frac{a_k(v_n) - a_k(u_n)}{v_n - u_n} = \sum_{k=0}^{n-1} a_k^-(x),$$

where a_k^- is the left-hand derivative. Since $a_k^-(x) = \pm 1$ for all x and k , the difference ratio cannot have a finite limit.

31.22. Let A be the x -set where (31.35) fails if f is replaced by $f\varphi$; then A has Lebesgue measure 0. Let G be the union of all open sets of μ -measure 0; represent G as a countable disjoint union of open intervals, and let B be G together with any endpoints of zero μ -measure of these intervals. Let D be the set of discontinuity points of F . If $F(x) \notin A$, $x \notin B$, and $x \notin D$, then $F(x-h) < F(x) < F(x+h)$, $F(x \pm h) \rightarrow F(x)$, and

$$\frac{1}{F(x+h) - F(x-h)} \int_{F(x-h)}^{F(x+h)} f(\varphi(t)) dt \rightarrow f(\varphi(F(x))).$$

Now $x - \epsilon < \varphi(F(x)) \leq x$ follows from $F(x - \epsilon) < F(x)$, and hence $\varphi(F(x)) = x$. If λ is Lebesgue measure restricted to $(0, 1)$, then $\mu = \lambda\varphi^{-1}$, and (31.36) follows by change of variable. But (36.36) is easy if $x \in D$, and hence it holds outside $B \cup (D^c \cap F^{-1}A)$. But $\mu(B) = 0$ by construction and $\mu(D^c \cap F^{-1}A) = 0$ by Problem 14.4.

Section 32

32.7. Define μ_n and ν_n as in (32.7), and write $\nu_n = \nu_{ac}^{(n)} + \nu_s^{(n)}$, where $\nu_{ac}^{(n)}$ is absolutely continuous with respect to μ_n and $\nu_s^{(n)}$ is singular with respect to μ_n . Take $\nu_{ac} = \sum_n \nu_{ac}^{(n)}$ and $\nu_s = \sum_n \nu_s^{(n)}$.

Suppose that $\nu_{ac}(E) + \nu_s(E) = \nu'_{ac}(E) + \nu'_s(E)$ for all E in \mathcal{F} . Choose an S in \mathcal{F} that supports ν_s and ν'_s and satisfies $\mu(S) = 0$. Then $\nu_{ac}(E) = \nu_{ac}(E \cap S^c) = \nu_{ac}(E \cap S^c) + \nu_s(E \cap S^c) = \nu'_{ac}(E \cap S^c) + \nu'_s(E \cap S^c) = \nu'_{ac}(E \cap S^c) = \nu'_{ac}(E)$. A similar argument shows that $\nu_s(E) = \nu'_s(E)$.

32.8. (a) Show that \mathcal{B} is closed under the formation of countable unions, choose \mathcal{B} -sets B_n such that $\mu(B_n) \rightarrow \sup_{\mathcal{B}} \mu(B) (< \infty)$, and take $B_0 = \bigcup_n B_n$.

(b) The same argument.

(c) Suppose $\mu(D_0) > 0$. The maximality of B_0 implies that $B_0 \cup D_0$ contains an E such that $\mu(E) > 0$ and $\nu(E) < \infty$. Since $B_0 \cap E \subset B_0 \in \mathcal{B}$, $\mu(B_0 \cap E) = 0$ ($\nu(E) < \infty$ rules out $\nu(B_0 \cap E) = \infty$). Therefore, $\mu(D_0 \cap E) > 0$ and $\nu(D_0 \cap E) < \infty$, which contradicts the maximality of C_0 .

(d) Take the density to be ∞ on Ω_0^c .

32.9. Define f and ν_s as in (32.8), and let f° and ν_s° be the corresponding function and measure for \mathcal{F}° : $\nu(E) = \int_E f^\circ d\mu + \nu_s^\circ(E)$ for $E \in \mathcal{F}^\circ$, and there is an \mathcal{F}° -set S° such that $\nu_s^\circ(\Omega - S^\circ) = 0$ and $\mu(S^\circ) = 0$. If $E \in \mathcal{F}^\circ$, it follows that $\int_E f^\circ d\mu = \int_{E-S^\circ} f^\circ d\mu = \int_{E-S^\circ} f^\circ d\mu^\circ = \nu^\circ(E - S^\circ) = \nu(E - S^\circ) \geq \int_{E-S^\circ} f d\mu = \int_E f d\mu$.

It is instructive to consider the extreme case $\mathcal{F}^\circ = \{0, \Omega\}$, in which ν° is absolutely continuous with respect to μ° (provided $\mu(\Omega) > 0$) and hence ν_s° vanishes.

Section 33

- 33.2.** (a) To prove independence, check the covariance. Now use Example 33.7.
 (b) Use the fact that R and Θ are independent (Example 20.2).
 (c) As the single event $[X = Y] = [X - Y = 0] = [\Theta = \pi/4] \cup [\Theta = 5\pi/4]$ has probability 0, the conditional probabilities have no meaning, and strictly speaking there is nothing to resolve. But whether it is natural to regard the degrees of freedom as one or as two depends on whether the 45° line through the origin is regarded as an element of the decomposition of the plane into 45° lines or whether it is regarded as the union of two elements of the decomposition of the plane into rays from the origin.
- Borel's paradox can be explained the same way: The equator is an element of the decomposition of the sphere into lines of constant latitude; the Greenwich meridian is an element of the decomposition of the sphere into great circles with common poles. The decomposition matters, which is to say the σ -field matters.
- 33.3.** (a) If the guard says, "1 is to be executed," then the conditional probability that 3 is also to be executed is $1/(1+p)$. The "paradox" comes from assuming that p must be 1, in which case the conditional probability is indeed $\frac{1}{2}$. But if $p \neq \frac{1}{2}$, then the guard does give prisoner 3 some information.
 (b) Here "one" and "other" are undefined, and the problem ignores the possibility that you have been introduced to a girl. Let the sample space be

$$\begin{array}{cccc} bbo \frac{\alpha}{4}, & bgo \frac{\beta}{4}, & gbo \frac{\gamma}{4}, & ggo \frac{\delta}{4}, \\ bby \frac{1-\alpha}{4}, & bgy \frac{1-\beta}{4}, & gby \frac{1-\gamma}{4}, & ggy \frac{1-\delta}{4}. \end{array}$$

For example, bgo is the event (probability $\beta/4$) that the older child is a boy, the younger is a girl, and the child you have been introduced to is the older; and ggy is the event (probability $(1-\delta)/4$) that both children are girls and the one you have been introduced to is the younger. Note that the four sex distributions do have probability $\frac{1}{4}$. If the child you have been introduced to is a boy, then the conditional probability that the other child is also a boy is $p = 1/(2 + \beta - \gamma)$. If $\beta = 1$ and $\gamma = 0$ (the parents present a son if they have one), then $p = \frac{1}{3}$. If $\beta = \gamma$ (the parents are indifferent), then $p = \frac{1}{2}$. Any p between $\frac{1}{3}$ and 1 is possible.

This problem shows again that one must keep in mind the entire experiment the sub- σ -field \mathcal{G} represents, not just one of the possible outcomes of the experiment.

- 33.6.** There is no problem, unless the notation gives rise to the illusion that $p(A|x)$ is $P(A \cap [X = x])/P[X = x]$.

- 33.15.** If N is a standard normal variable, then

$$\frac{1}{\sqrt{n}} p_y \left(y + \frac{x}{\sqrt{n}} \right) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2} f \left(y + \frac{x}{\sqrt{n}} \right) \bigg/ E \left[f \left(y + \frac{N}{\sqrt{n}} \right) \right].$$

Section 34

34.3. If (X, Y) takes the values $(0, 0)$, $(1, -1)$, and $(1, 1)$ with probability $\frac{1}{3}$ each, then X and Y are dependent but $E[Y|X] = E[Y] = 0$.

If (X, Y) takes the values $(-1, 1)$, $(0, -2)$, and $(1, 1)$ with probability $\frac{1}{3}$ each, then $E[X] = E[Y] = E[XY] = 0$ and so $E[XY] = E[X]E[Y]$, but $E[Y|X] = Y \neq 0 = E[Y]$. Of course, this is another example of dependent but uncorrelated random variables.

34.4. First show that $\int f dP_0 = \int_B f dP / P(B)$ and that $P[B|\mathcal{G}] > 0$ on a set of P_0 -measure 1. Let G be the general set in \mathcal{G} .

(a) Since

$$\begin{aligned} \int_G P_0[A|\mathcal{G}] P[B|\mathcal{G}] dP &= \int_G P_0[A|\mathcal{G}] I_B dP = \int_B I_G P_0[A|\mathcal{G}] dP \\ &= P(B) \int_\Omega I_G P_0[A|\mathcal{G}] dP_0 = P(B) P_0(A \cap G) \\ &= \int_G P[A \cap B|\mathcal{G}] dP, \end{aligned}$$

it follows that

$$P_0[A|\mathcal{G}] P[B|\mathcal{G}] = P[A \cap B|\mathcal{G}]$$

holds on a set of P -measure 1.

(b) If $P_i(A) = P(A|B_i)$, then

$$\begin{aligned} \int_{G \cap B_i} P_i[A|\mathcal{G}] dP &= P(B_i) \int_\Omega I_G P_i[A|\mathcal{G}] dP_i = P(B_i) P_i(A \cap G) \\ &= \int_{G \cap B_i} P[A|\mathcal{G} \vee \mathcal{H}] dP. \end{aligned}$$

Therefore, $\int_C I_{B_i} P_i[A|\mathcal{G}] dP = \int_C I_{B_i} P[A|\mathcal{G} \vee \mathcal{H}] dP$ if $C = G \cap B_i$, and of course this holds for $C = G \cap B_j$ if $j \neq i$. But C 's of this form constitute a π -system generating $\mathcal{G} \vee \mathcal{H}$, and hence $I_{B_i} P_i[A|\mathcal{G}] = I_{B_i} P[A|\mathcal{G} \vee \mathcal{H}]$ on a set of P -measure 1. Now use the result in part (a).

34.9. All such results can be proved by imitating the proofs for the unconditional case or else by using Theorem 34.5 (for part (c), as generalized in Problem 34.7). For part (a), it must be shown that it is possible to take the integral measurable \mathcal{G} .

34.10. (a) If $Y = X - E[X|\mathcal{G}_1]$, then $X - E[X|\mathcal{G}_2] = Y - E[Y|\mathcal{G}_2]$, and $E[(Y - E[Y|\mathcal{G}_2])^2|\mathcal{G}_2] = E[Y^2|\mathcal{G}_2] - E^2[Y|\mathcal{G}_2] \leq E[Y^2|\mathcal{G}_2]$. Take expected values.

34.11. First prove that

$$P[A_1 \cap A_3|\mathcal{G}_2] = E[I_{A_1} P[A_3|\mathcal{G}_{12}]|\mathcal{G}_2].$$

From this and (i) deduce (ii). From

$$E[I_{A_1} P[A_3 \| \mathcal{G}_2] | \mathcal{G}_2] = P[A_1 | \mathcal{G}_2] P[A_3 | \mathcal{G}_2],$$

(ii), and the preceding equation deduce

$$\int_{A_1 \cap A_2} P[A_3 | \mathcal{G}_2] dP = \int_{A_1 \cap A_2} P[A_3 | \mathcal{G}_{12}] dP.$$

The sets $A_1 \cap A_2$ form a π -system generating \mathcal{G}_{12} .

34.16. (a) Obviously (34.18) implies (34.17). If (34.17) holds, then clearly (34.18) holds for X simple. For the general X , choose simple X_k such that $\lim_k X_k = X$ and $|X_k| \leq |X|$. Note that

$$\begin{aligned} & \left| \int_{A_n} X dP - \alpha \int X dP \right| \\ & \leq \left| \int_{A_n} X_k dP - \alpha \int X_k dP \right| + (1 + |\alpha|) E[|X - X_k|]; \end{aligned}$$

let $n \rightarrow \infty$ and then let $k \rightarrow \infty$.

(b) If $\Omega \in \mathcal{P}$, then the class of E satisfying (34.17) is a λ -system, and so by the π - λ theorem and part (a), (34.18) holds if X is measurable $\sigma(\mathcal{P})$. Since $A_n \in \sigma(\mathcal{P})$, it follows that

$$\begin{aligned} \int_{A_n} X dP &= \int_{A_n} E[X | \sigma(\mathcal{P})] dP \rightarrow \alpha \int E[X | \sigma(\mathcal{P})] dP \\ &= \alpha \int X dP. \end{aligned}$$

(c) Replace X by $X dP_0/dP$ in (34.18).

34.17. (a) The Lindeberg–Lévy theorem.

(b) Chebyshev's inequality.

(c) Theorem 25.4.

(d) Independence of the X_n .

(e) Problem 34.16(b).

(f) Problem 34.16(c).

(g) Part (b) here and the ϵ - δ definition of absolute continuity.

(h) Theorem 25.4 again.

Section 35

35.4. (b) Let S_n be the number of k such that $1 \leq k \leq n$ and $Y_k = \frac{3}{2}$. Then $X_n = 3^{S_n}/2^n$. Take logarithms and use the strong law of large numbers.

35.9. Let K bound $|X_1|$ and the $|X_n - X_{n-1}|$. Bound $|X_\tau|$ by $K\tau$. Write $\int_{\tau \leq k} X_\tau dP = \sum_{i=1}^k \int_{\tau=i} X_i dP = \sum_{i=1}^k (\int_{\tau \geq i} X_i dP - \int_{\tau \geq i+1} X_i dP)$. Transform the last integral by the martingale property and reduce the expression to $E[X_1] - \int_{\tau > k} X_{k+1} dP$. Now

$$\left| \int_{\tau > k} X_{k+1} dP \right| \leq K(k+1)P[\tau > k] \leq K(k+1)k^{-1} \int_{\tau > k} \tau dP \rightarrow 0.$$

35.13. (a) By the result in Problem 32.9, X_1, X_2, \dots is a supermartingale. Since $E[|X_n|] = E[X_n] \leq \nu(\Omega)$, Theorem 35.5 applies.

(b) If $A \in \mathcal{F}_n$, then $\int_A (Y_n + Z_n) dP + \sigma'_n(A) = \int_A X_\infty dP + \sigma_\infty(A) = \nu(A) = \int_A X_n dP + \sigma_n(A)$. Since the Lebesgue decomposition is unique (Problem 32.7), $Y_n + Z_n = X_n$ with probability 1. Since X_n and Y_n converge, so does Z_n . If $A \in \mathcal{F}_k$ and $n \geq k$, then $\int_A Z_n dP \leq \sigma_\infty(A)$, and by Fatou's lemma, the limit Z satisfies $\int_A Z dP \leq \sigma_\infty(A)$. This holds for A in $\bigcup_k \mathcal{F}_k$ and hence (monotone class theorem) for A in \mathcal{F}_∞ . Choose A so that $P(A) = 1$ and $\sigma_\infty(A) = 0$: $E[Z] = \int_A Z dP \leq \sigma_\infty(A) = 0$.

It can happen that $\sigma_n(\Omega) = 0$ and $\sigma_\infty(\Omega) = \nu(\Omega) > 0$, in which case σ_n does not converge to σ_∞ and the X_n cannot be integrated to the limit.

35.17. For a very general result, see J. L. Doob: Application of the theory of martingales, *Le Calcul des Probabilités et ses Applications* (Colloques Internationaux du Centre de la Recherche Scientifique, Paris, 1949).

Section 36

36.5. (b) Show by part (a) and Problem 34.18 that f_n is the conditional expected value of f with respect to the σ -field \mathcal{T}_{n+1} generated by the coordinates x_{n+1}, x_{n+2}, \dots . By Theorem 35.9, (36.30) will follow if each set in $\bigcap_n \mathcal{T}_n$ has π -measure either 0 or 1, and here the zero-one law applies.

(c) Show that g_n is the conditional expected value of f with respect to the σ -field generated by the coordinates x_1, \dots, x_n , and apply Theorem 35.6.

36.7. Let \mathcal{L} be the countable set of simple functions $\sum_i \alpha_i I_{A_i}$ for α_i rational and $\{A_i\}$ a finite decomposition of the unit interval into subintervals with rational endpoints. Suppose that the X_t exist, and choose (Theorem 17.1) Y_t in \mathcal{L} so that $E[|X_t - Y_t|] < \frac{1}{4}$. From $E[|X_s - X_t|] = \frac{1}{2}$, conclude that $E[|Y_s - Y_t|] > 0$ for $s \neq t$. But there are only countably many of the Y_t . It does no good to replace Lebesgue measure by some other measure on the unit interval.

Section 37

37.1. If t_1, \dots, t_k are in increasing order and $t_0 = 0$, then

$$\begin{aligned} \sum_{i,j} K(t_i, t_j) x_i x_j &= \sum_{i,j} x_i x_j \sum_{l=1}^{\min(i,j)} (t_l - t_{l-1}) \\ &= \sum_l (t_l - t_{l-1}) \left(\sum_{i \geq l} x_i \right)^2 \geq 0. \end{aligned}$$

37.4 (a) Use Problem 36.6(b).

(b) Let $[W_t: t \geq 0]$ be a Brownian motion on $(\Omega, \mathcal{F}, P_0)$, where $W(\cdot, \omega) \in C$ for every ω . Define $\xi: \Omega \rightarrow R^T$ by $Z_t(\xi(\omega)) = W_t(\omega)$. Show that ξ is measurable $\mathcal{F}/\mathcal{R}^T$ and $P = P_0 \xi^{-1}$. If $C \subset A \in \mathcal{R}^T$, then $P(A) = P_0(\xi^{-1}A) = P_0(\Omega) = 1$.

37.5. Consider $W(1) = \sum_{k=1}^n (W(k/n) - W((k-1)/n))$ for notational convenience. Since

$$n \int_{|W(1/n)| \geq \epsilon} W^2\left(\frac{1}{n}\right) dP = \int_{|W(1)| \geq \epsilon \sqrt{n}} W^2(1) dP \rightarrow 0,$$

the Lindeberg theorem applies.

37.14. By symmetry,

$$\rho(s, t) = 2P\left[W_s > 0, \inf_{s \leq u \leq t} (W_u - W_s) \leq -W_s\right];$$

W_s and the infimum here are independent because of the Markov property, and so by (20.30) (and symmetry again)

$$\begin{aligned} \rho(s, t) &= 2 \int_0^\infty P[\tau_x \leq t-s] \frac{1}{\sqrt{2\pi s}} e^{-x^2/2s} dx \\ &= 2 \int_0^\infty \int_0^{t-s} \frac{x}{\sqrt{2\pi}} \frac{1}{u^{3/2}} e^{-x^2/2u} \frac{1}{\sqrt{2\pi s}} e^{-x^2/2s} du dx. \end{aligned}$$

Reverse the integral, use $\int_0^\infty x e^{-x^2 r/2} dx = 1/r$, and put $v = (s/(s+u))^{1/2}$:

$$\begin{aligned} \rho(s, t) &= \frac{1}{\pi} \int_0^{t-s} \frac{1}{u+s} \frac{s^{1/2}}{u^{1/2}} du \\ &= \frac{2}{\pi} \int_{\sqrt{s/t}}^1 \frac{dv}{\sqrt{1-v^2}}. \end{aligned}$$

Bibliography

HALMOS and SAKS have been the strongest measure-theoretic and DOOB and FELLER the strongest probabilistic influences on this book, and the spirit of KAC's small volume has been very important.

AUBREY: *Brief Lives*, John Aubrey; ed., O. L. Dick. Seker and Warburg, London, 1949.

BAHADUR: *Some Limit Theorems in Statistics*, R. R. Bahadur. SIAM, Philadelphia, 1971.

BANACH: *Théorie des Opérations Linéaires*, S. Banach. Monografie Matematyczne, Warsaw, 1932.

BERGER: *Statistical Decision Theory*, 2nd ed., James O. Berger. Springer-Verlag, New York, 1985.

BHATTACHARYA & WAYMIRE: *Stochastic Processes with Applications*, Rabi N. Bhattacharya and Edward C. Waymire. Wiley, New York, 1990.

BILLINGSLEY₁: *Convergence of Probability Measures*, Patrick Billingsley. Wiley, New York, 1968.

BILLINGSLEY₂: *Weak Convergence of Measures: Applications in Probability*, Patrick Billingsley. SIAM, Philadelphia, 1971.

BIRKHOFF & MAC LANE: *A Survey of Modern Algebra*, 4th ed., Garrett Birkhoff and Saunders Mac Lane. Macmillan, New York, 1977.

ÇINLAR: *Introduction to Stochastic Processes*, Erhan Çinlar. Prentice-Hall, Englewood Cliffs, New Jersey, 1975.

CRAMÉR: *Mathematical Methods of Statistics*, Harald Cramér. Princeton University Press, Princeton, New Jersey, 1946.

DOOB: *Stochastic Processes*, J. L. Doob. Wiley, New York, 1953.

DUBINS & SAVAGE: *How to Gamble If You Must*, Lester E. Dubins and Leonard J. Savage. McGraw-Hill, New York, 1965.

- DUDLEY: *Real Analysis and Probability*, Richard M. Dudley. Wadsworth and Brooks, Pacific Grove, California, 1989.
- DYNKIN & YUSHKEVICH: *Markov Processes*, English ed., Evgenii B. Dynkin and Aleksandr A. Yushkevich. Plenum Press, New York, 1969.
- FELLER: *An Introduction to Probability Theory and Its Applications*, Vol. I. 3rd ed., Vol. II, 2nd ed., William Feller. Wiley, New York, 1968, 1971.
- GALAMBOS: *The Asymptotic Theory of Extreme Order Statistics*, Janos Galambos. Wiley, New York, 1978.
- GELBAUM & OLMSTED: *Counterexamples in Analysis*, Bernard R. Gelbaum and John M. Olmsted. Holden-Day, San Francisco, 1964.
- GNEDENKO & KOLMOGOROV: *Limit Distributions for Sums of Independent Random Variables*, English ed., B. V. Gnedenko and A. N. Kolmogorov. Addison-Wesley, Reading, Massachusetts, 1954.
- HALMOS₁: *Measure Theory*, Paul R. Halmos. Van Nostrand, New York, 1950.
- HALMOS₂: *Naive Set Theory*, Paul R. Halmos. Van Nostrand, Princeton, 1960.
- HARDY: *A Course of Pure Mathematics*, 9th ed., G. H. Hardy. Macmillan, New York, 1946.
- HARDY & WRIGHT: *An Introduction to the Theory of Numbers*, 4th ed., G. H. Hardy and E. M. Wright. Clarendon, Oxford, 1959.
- HAUSDORFF: *Set Theory*, 2nd English ed., Felix Hausdorff. Chelsea, New York, 1962.
- JECH: *Set Theory*, Thomas Jech. Academic Press, New York, 1978.
- KAC: *Statistical Independence in Probability, Analysis and Number Theory*, Carus Math. Monogr. 12, Marc Kac. Wiley, New York, 1959.
- KAHANE: *Some Random Series of Functions*, Jean-Pierre Kahane. Heath, Lexington, Massachusetts, 1968.
- KAPLANSKY: *Set Theory and Metric Spaces*, Irving Kaplansky. Chelsea, New York, 1972.
- KARATZES & SHREVE: *Brownian Motion and Stochastic Calculus*, Ioannis Karatzis and Steven E. Shreve. Springer-Verlag, New York, 1988.
- KARLIN & TAYLOR: *A First Course in Stochastic Processes*, 2nd ed., *A Second Course in Stochastic Processes*, Samuel Karlin and Howard M. Taylor. Academic Press, New York, 1975 and 1981.
- KOLMOGOROV: *Grundbegriffe der Wahrscheinlichkeitsrechnung*, Erg. Math., Vol. 2, No. 3, A. N. Kolmogorov. Springer-Verlag, Berlin, 1933.
- LÉVY: *Théorie de l'Addition des Variables Aléatoires*, Paul Lévy. Gauthier-Villars, Paris, 1937.
- PROTTER: *Stochastic Integration and Differential Equations*, Philip Protter. Springer-Verlag, 1990.

- RIESZ & SZ.-NAGY: *Functional Analysis*, English ed., Frigyes Riesz and Bela Sz.-Nagy. Unger, New York, 1955.
- ROCKETT & SZÜSZ: *Continued Fractions*, Andrew M. Rockett and Peter Szüsz. World Scientific, Singapore, 1992.
- ROYDEN: *Real Analysis*, 2nd ed., H. I. Royden. Macmillan, New York, 1968.
- RUDIN₁: *Principles of Mathematical Analysis*, 3rd ed., Walter Rudin. McGraw-Hill, New York, 1976.
- RUDIN₂: *Real and Complex Analysis*, 2nd ed., Walter Rudin. McGraw-Hill, New York, 1974.
- SAKS: *Theory of the Integral*, 2nd rev. ed., Stanislaw Saks. Hafner, New York, 1937.
- SPIVAK: *Calculus on Manifolds*, Michael Spivak. W. A. Benjamin, New York, 1965.
- WAGON: *The Banach–Tarsky Paradox*, Stan Wagon. Cambridge University Press, 1985.

List of Symbols

| | |
|---------------------------------|--|
| Ω , 536 | $P(B A)$, 51 |
| 2^Ω , 536 | $\limsup_n A_n$, 52 |
| \emptyset , 536 | $\liminf_n A_n$, 52 |
| $(a, b]$, 537 | $\lim_n A_n$, 52 |
| I_A , 536 | i.o., 53 |
| $ I $, 1 | \mathcal{T} , 62, 287 |
| P , 2, 22 | R^1 , 537 |
| $d_n(\omega)$, 3 | R^k , 539 |
| $r_n(\omega)$, 5 | $[X = x]$, 67 |
| $s_n(\omega)$, 6 | $\sigma(X)$, 68, 255 |
| N , 8, 357 | μ , 73, 160, 256 |
| $\lfloor x \rfloor$, 537 | $E[X]$, 76, 273 |
| $\operatorname{sgn} x$, 537 | $\operatorname{Var}[X]$, 78, 275 |
| $A - B$, 536 | $E_n[f]$, 87 |
| A^c , 536 | $s_c(a)$, 93 |
| $A \triangle B$, 536 | τ , 99, 133, 464, 508 |
| $A \subset B$, 536 | p_{ij} , 111 |
| \mathcal{F} -set, 20 | S , 111 |
| \mathcal{B}_0 , 20 | α_i , 111 |
| $\sigma(\mathcal{A})$, 21 | π_i , 124 |
| \mathcal{I} , 22 | $M(t)$, 146, 278, 285 |
| \mathcal{B} , 22 | \mathcal{R}^k , 158 |
| (Ω, \mathcal{F}, P) , 23 | $\mathcal{A} \cap \Omega_0$, 159 |
| $A_n \uparrow A$, 536 | $x_k \uparrow x$, 160, 537 |
| $A_n \downarrow A$, 536 | $x = \sum_k x_k$, 160 |
| λ , 25, 43, 168 | μ^* , 165 |
| \wedge , 537 | $\mathcal{H}(\mu^*)$, 165 |
| \vee , 537 | λ_1 , 168 |
| $f(\mathcal{A})$, 33 | λ_k , 171 |
| $P_n(A)$, 35 | F , 175, 177 |
| $D(A)$, 35 | $\Delta_A F$, 176 |
| \mathcal{D} , 35 | $T^{-1}A'$, 537 |
| P^* , 37, 47 | \mathcal{F}/\mathcal{F}' , 182 |
| P_* , 37, 47 | $F_n \Rightarrow F$, 191, 327, 378 |
| S^∞ , 27, 311 | $dF(x)$, 228 |
| \mathcal{P} , 41 | $X \times Y$, 231 |
| \mathcal{L} , 41 | $\mathcal{X} \times \mathcal{Y}$, 231 |
| λ^* , 44 | $\mu \times \nu$, 233 |

$*$, 266

$X_n \rightarrow_p X$, 70, 268, 330

$\|f\|$, 249

$\|f\|_p$, 241

L^p , 241

$\mu_n \Rightarrow \mu$, 327, 378

$X_n \Rightarrow X$, 329, 378

$X_n \Rightarrow a$, 331

∂A , 538

$\varphi(t)$, 342

$\mu_n \rightarrow_i \mu$, 371

F_s , 414

F_{ac} , 414

$\nu \ll \mu$, 422

$d\nu/d\mu$, 423

ν_s , 424

ν_{ac} , 424

$P[A|\mathcal{G}]$, 428, 430

$P[A|X_t, t \in T]$, 433

$E[X|\mathcal{G}]$, 445

R^T , 484

\mathcal{R}^T , 485

W_t , 498

Index

Here A_n refers to paragraph n of the Appendix (p. 536); $u.v$ refers to Problem v in Section u , or else to a note on it (Notes on the Problems, p. 552); the other references are to pages. Greek letters are alphabetized by their Roman equivalents (m for μ , and so on). Names in the bibliography are not indexed separately.

- Absolute continuity, 413, 422
- Absolutely continuous part, 425
- Absolute moment, 274
- Absorbing state, 112
- Adapted σ -fields, 458
- Additive set function, 420
- Additivity:
 - countable, 23, 161
 - finite, 23, 161
- Admissible, 248, 252
- Affine transformation, 172
- Algebra, 19
- Almost everywhere, 60
- Almost surely, 60
- α -mixing, 363, 29.10
- Aperiodic, 125
- Approximation of measure, 168
- Area over the curve, 79
- Area under the curve, 203
- Asymptotic equipartition property, 91, 144
- Asymptotic relative frequency, 8
- Atom, 271
- Autoregression, 495
- Axiom of choice, 21, 45

- Baire category, 1.10, A15
- Baire function, 13.7
- Banach limits, 3.8, 19.3
- Banach space, 243
- Banach–Tarski paradox, 180
- Bayes estimation, 475
- Bayes risk, 248, 251
- Benford law, 25.3

- Beppo Levi theorem, 16.3
- Bernoulli–Laplace model of diffusion, 112
- Bernoulli shift, 311
- Bernoulli trials, 75
- Bernstein polynomial, 87
- Betting system, 98
- Binary digit, 3
- Binomial distribution, 256
- Blackwell–Rao theorem, 455
- Bold play, 102
- Boole’s inequality, 25
- Borel, 9
- Borel–Cantelli lemmas, 59, 60
- Borel function, 183
- Borel normal number theorem, 9
- Borel paradox, 441
- Borel set, 22, 158
- Boundary, A11
- Bounded convergence theorem, 210
- Bounded variation, 415
- Branching process, 461
- Britannica, 552
- Brownian motion, 498
- Burstin’s theorem, 22.14

- Canonical measure, 372
- Canonical representation, 372
- Cantelli inequality, 5.5
- Cantelli theorem, 6.6
- Cantor function, 31.2, 31.15
- Cantor set, 1.5
- Cardinality of σ -fields, 2.12, 2.22
- Cartesian product, 231

- Category, A15, 1.10
- Cauchy distribution, 20.14, 348
- Cauchy equation, A20, 14.7
- Cavalieri principle, 18.8
- Central limit theorem, 291, 357, 385, 391, 34.17, 475
- Cesàro averages, A30, 20.23
- Change of variable, 215, 224, 225, 274
- Characteristic function, 342
- Chebyshev inequality, 5, 80, 276
- Chernoff theorem, 151
- Chi-squared distribution, 20.15
- Chi-squared statistic, 29.8
- Circular Lebesgue measure, 13.12, 313
- Class of sets, 18
- Closed set, A11
- Closed set of states, 8.21
- Closed support, 12.9
- Closure, A11
- Cocountable set, 21
- Cofinite set, 20
- Collective, 109
- Compact, A13
- Complement, A1
- Completely normal number, 6.13
- Complete space or measure, 44, 10.5
- Completion, 3.10, 10.5
- Complex functions, integration of, 218
- Compound Poisson distribution, 28.3
- Compound Poisson process, 32.7
- Concentrated, 161
- Conditional distribution, 439, 449
- Conditional expected value, 133, 445
- Conditional probability, 51, 427, 33.5
- Congruent by dissection, 179
- Conjugate index, 242
- Conjugate space, 244
- Consistency conditions for finite-dimensional distributions, 483
- Content, 3.15
- Continued-fraction transformation, 319, A36
- Continuity from above, 25
- Continuity of paths, 500
- Continuum hypothesis, 46
- Conventions involving ∞ , 160
- Convergence in distribution, 329, 378
- Convergence in mean, 243
- Convergence in measure, 268
- Convergence in probability, 70, 268, 330
- Convergence with probability 1, 70, 330
- Convergence of random series, 289
- Convergence of types, 193
- Convex functions, A32
- Convolution, 266
- Coordinate function, 27, 484
- Coordinate variable, 484
- Countable, 8
- Countable additivity, 23, 161
- Countable subadditivity, 25, 162
- Countably generated σ -field, 2.11
- Countably infinite, 8
- Counting measure, 161
- Coupled chain, 126
- Coupon problem, 362
- Covariance, 277
- Cover, A3
- Cramér-Wold theorem, 383
- Cylinder, 27, 485
- Daniell–Stone theorem, 11.14, 16.12
- Darboux–Young definition, 15.2
- Δ -distribution, 192
- Decision theory, 247
- Decomposition, A3
- de Finetti theorem, 473
- Definite integral, 200
- Degenerate distribution function, 193
- Delta method, 359
- DeMoivre–Laplace theorem, 25.11, 358
- DeMorgan law, A6
- Dense, A15
- Density of measure, 213, 422
- Density point, 31.9
- Density of random variable or distribution, 257, 260
- Density of set of integers, 2.18
- Denumerable probabilities, 51
- Dependent random variables, 363
- Derivatives of integrals, 402
- Diagonal method, 29, A14
- Difference equation, A19
- Difference set, A1
- Diophantine approximation, 13, 324
- Dirichlet theorem, 13, A26
- Discontinuity of the first kind, 534
- Discrete measure, 23, 161
- Discrete random variable, 256
- Discrete space, 1.1, 23, 5.16
- Disjoint, A3
- Disjoint supports, 410, 421
- Distribution:
 - of random variable, 73, 187, 256
 - of random vector, 259
- Distribution function, 175, 188, 256, 259, 409
- Dominated convergence theorem, 78, 209
- Dominated measure, 422
- Double exponential distribution, 348
- Double integral, 233

- Double series, A27
- Doubly stochastic matrix, 8.20
- Dual space, 245
- Dubins–Savage theorem, 102
- Dyadic expansion, 3, A31
- Dyadic interval, 4
- Dyadic transformation, 313
- Dynkin's π – λ theorem, 42
- ϵ – δ definition of absolute continuity, 422
- Egorov theorem, 13.9
- Eigenvalues, 8.26
- Empirical distribution function, 268
- Empty set, A1
- Entropy, 57, 6.14, 8.31, 31.17
- Equicontinuous, 355
- Equivalence class, 58
- Equivalent measures, 422
- Erdős–Kac central limit theorem, 395
- Ergodic theorem, 314
- Erlang density, 23.2
- Essential supremum, 241
- Estimation, 251, 452
- Etemadi, 282, 288, 22.15
- Euclidean distance, A16
- Euclidean space, A1, A16
- Euler function, 2.18
- Event, 18
- Excessive function, 134
- Exchangeable, 473
- Existence of independent sequences, 73, 265
- Existence of Markov chains, 115
- Expected value, 76, 273
- Exponential convergence, 131, 8.18
- Exponential distribution, 189, 258, 297, 348
- Extension of measure, 36, 166, 11.1
- Extremal distribution, 195
- Factorization and sufficiency, 450
- Fair game, 92, 463
- Fatou lemma, 209
- Field, 19, 2.5
- Filtration, 458
- Finite additivity, 20, 23, 2.15, 3.8, 161
- Finite or countable, 8
- Finite-dimensional distributions, 308, 482
- Finite-dimensional sets, 485
- Finitely additive field, 20
- Finite subadditivity, 24, 162
- First Borel–Cantelli lemma, 59
- First category, 1.10, A15
- First passage, 118
- fixed discontinuity, 303
- Fourier representation, 250
- Fourier series, 351, 26.30
- Fourier transform, 342
- Frequency, 8
- Fubini theorem, 233
- Functional central limit theorem, 522
- Fundamental in probability, 20.21
- Fundamental set, 320
- Fundamental theorem of calculus, 224, 400
- Fundamental theorem of Diophantine approximation, 324
- Gambling policy, 98
- Gamma distribution, 20.17
- Gamma function, 18.18
- Generated σ -field, 21
- Glivenko–Cantelli theorem, 269
- Goncharov's theorem, 361
- Hahn decomposition, 420
- Hamel basis, 14.7
- Hardy–Ramanujan theorem, 6.16
- Heine–Borel theorem, A13, A17
- Hewitt–Savage zero-one law, 496
- Hilbert space, 249
- Hitting time, 136
- Hölder's inequality, 80, 5.9, 242, 276
- Hypothesis testing, 151
- Inadequacy of \mathcal{R}^T , 492
- Inclusion-exclusion formula, 24, 163
- Indefinite integral, 400
- Identically distributed, 85
- Independent classes, 55
- Independent events, 53
- Independent increments, 299, 498
- Independent random variables, 71, 261
- Independent random vectors, 263
- Indicator, A5
- Infinitely divisible distributions, 371
- Infinitely often, 53
- Infinite series, A25
- Information, 57
- Initial digit problem, 25.3
- Initial probabilities, 111
- Inner boundary, 64
- Inner measure, 37, 3.2
- Integrable, 200, 206
- Integral, 199
- Integral with respect to Lebesgue measure, 221
- Integrals of derivatives, 412
- Integration by parts, 236
- Integration over sets, 212
- Integration with respect to a density, 214

- Interior, A11
- Interval, A9
- Invariance principle, 520
- Invariant set, 313
- Inverse image, A7, 182
- Inversion formula, 346
- Irreducible chain, 119
- Irregular paths, 504
- Iterated integral, 233
- Jacobian, 225, 261, 545
- Jensen inequality, 80, 276, 449
- Jordan decomposition, 421
- Jordan measurable, 3.15
- k -dimensional Borel set, 158
- k -dimensional Lebesgue measure, 171, 177, 17.14, 20.4
- Kolmogorov existence theorem, 483
- Kolmogorov zero-one law, 63, 287
- Landau notation, A18
- Laplace distribution, 348
- Laplace transform, 285
- Large deviations, 148
- Lattice distribution, 26.1
- Law of the iterated logarithm, 153
- Law of large numbers:
 - strong, 9, 11, 85, 282
 - weak, 5, 11, 86, 284
- Lebesgue decomposition, 414, 425
- Lebesgue density theorem, 31.9, 35.15
- Lebesgue function, 31.3
- Lebesgue integrable, 221, 225
- Lebesgue measure, 25, 43, 167, 171, 177
- Lebesgue set, 45
- Leibniz formula, 17.8
- Lévy distance, 14.5, 25.4, 26.16
- Likelihood ratio, 461, 471
- Limit inferior, 52, 4.1
- Limit of sets, 52
- Lindeberg condition, 359
- Lindeberg–Lévy theorem, 357
- Linear Borel set, 158
- Linear functional, 244
- Linearity of expected value, 77
- Linearity of the integral, 206
- Linearly independent reals, 14.7, 30.8
- Lipschitz condition, 418
- Log-normal distribution, 388
- Lower integral, 204, 228
- Lower semicontinuous, 29.1
- Lower variation, 421
- L^p -space, 241
- λ -system, 41
- Lusin theorem, 17.10
- Lyapounov condition, 362
- Lyapounov inequality, 81, 277
- Mapping theorem, 344, 380
- Marginal distribution, 261
- Markov chain, 111, 363, 367, 29.11, 429
- Markov inequality, 80, 276
- Markov process, 435, 510
- Markov shift, 312
- Markov time, 133
- Martingale, 101, 458, 514
- Martingale central limit theorem, 475
- Martingale convergence theorem, 468
- Maximal ergodic theorem, 317
- Maximal inequality, 287
- Maximal solution, 122
- μ -continuity set, 335, 378
- m -dependent, 6.11, 364
- Mean value, 26.17
- Measurable mapping, 182
- Measurable process, 503
- Measurable rectangle, 231
- Measurable with respect to a σ -field, 68, 225
- Measurable set, 20, 38, 165
- Measurable space, 161
- Measure, 22, 160
- Measure-preserving transformation, 311
- Measure space, 161
- Meets, A3
- Method of moments, 388, 30.6
- Minimal sufficient field, 454
- Minimum-variance estimation, 454
- Minkowski inequality, 5.10, 242
- Mixing, 24.3, 363, 29.10, 34.16
- Mixture, 473
- μ^* -measurable, 165
- Moment, 274
- Moment generating function, 1.6, 146, 278, 284, 390
- Monotone, 24, 162, 206
- Monotone class, 43, 3.12
- Monotone class theorem, 43
- Monotone convergence theorem, 208
- M-test, 210, A28
- Multidimensional central limit theorem, 385
- Multidimensional characteristic function, 381
- Multidimensional distribution, 259
- Multidimensional normal distribution, 383
- Multinomial sampling, 29.8
- Negative part, 200, 254
- Negligible set, 8, 1.3, 1.9, 44

- Neyman–Pearson lemma, 19.7
- Nonatomic, 2.19
- Nondenumerable probabilities, 526
- Nonmeasurable set, 45, 12.4
- Nonnegative series, A25
- Norm, 243
- Normal distribution, 258, 383
- Normal number, 8, 1.8, 86, 6.13
- Normal number theorem, 9, 6.9
- Nowhere dense, A15
- Nowhere differentiable, 31.18, 505
- Null persistent, 130
- Number theory, 393

- Open set, A11
- Optional sampling theorem, 466
- Optimal stopping, 133
- Order of dyadic interval, 4
- Orthogonal projection, 250
- Orthonormal, 249
- Ottaviani inequality, 22.15
- Outer boundary, 64
- Outer measure, 37, 3.2, 165

- Pairwise disjoint, A3
- Partial-fraction expansion, 20.14
- Partial information, 57
- Partition, A3
- Path function, 308, 493, 500
- Payoff function, 133
- Perfect set, A15
- Peano curve, 179
- Period, 125
- Permutation, 72, 86, 361
- Persistent, 117, 120
- Phase space, 111
- π - λ theorem, 42
- P^* -measurable, 38
- Poincaré theorem, 29.9
- Point of increase, 12.9, 20.12
- Poisson approximation, 302, 328
- Poisson distribution, 257, 299, 375
- Poisson process, 297, 436
- Poisson theorem, 6.5
- Polar coordinates, 226, 261
- Polya criterion, 26.3
- Polya theorem, 118
- Positive part, 200, 254
- Positive persistent, 130
- Power class, 21, A1
- Power series, A29
- Prékopà theorem, 303
- Primitive, 224, 400
- Probability measure, 22
- Probability measure space, 23
- Probability transformation, 14.3
- Product measure, 28, 12.12, 233, 487
- Product space, 27, 231, 484
- Projection, 27, 484
- Proper difference, A1
- Proper subset, A1
- π -system, 41

- Rademacher functions, 5, 289, 291
- Radon–Nykodym derivative, 423, 460, 470
- Radon–Nykodym theorem, 422, 32.8
- Random Taylor series, 292
- Random variable, 67, 182, 254
- Random vector, 183, 255
- Random walk, 112
- Rank, 4, 320
- Ranks and records, 20.8
- Rate of Poisson process, 299
- Rational rectangle, 158
- Realization of process, 493
- Record values, 20.9, 21.3, 22.9, 27.8
- Rectangle, 158, A16
- Recurrent event, 8.17
- Red-and-black, 92
- Reflection principle, 511
- Regularity, 174
- Relative frequency, 8
- Relative measure, 25.16
- Renewal theory, 8.17, 310
- Reversed martingale, 472
- Riemann integral, 2, 12, 221, 228, 25.12
- Riemann–Lebesgue lemma, 345
- Riesz–Fischer theorem, 243
- Riesz representation theorem, 17.12, 244
- Right continuity, 175, 256
- Rigid transformation, 172
- Risk, 247, 251

- Saltus, 188
- Sample function, 188
- Sample path, 188, 308
- Sample point, 18
- Sampling theory, 392
- Scheffé theorem, 215
- Schwarz inequality, 81, 5.6, 249, 276
- Second Borel–Cantelli lemma, 60, 4.11, 88
- Second category, 1.10, A15
- Second-order Markov chain, 8.32
- Secretary problem, 114
- Section, 231
- Selection problem, 113, 138
- Selection system, 95, 7.3
- Semicontinuous, 29.1

- Semiring, 166
- Separable function, 526
- Separable process, 527, 531
- Separable σ -field, 2.11
- Separant, 527
- Sequence space, 27, 311
- Set function, 22, 420
- σ -field, 20
- σ -field generated by a class of sets, 21
- σ -field generated by a random variable, 68, 255
- σ -finite on a class, 160
- σ -finite measure, 160
- Shannon theorem, 6.14, 8.31
- Signed measure, 32.12
- Simple function, 184
- Simple random variable, 67, 185, 254
- Singleton, A1
- Singular function, 407
- Singularity, 421
- Singular part, 425
- Skorohod embedding, 513, 519
- Skorohod theorem, 333
- Southwest, 176, 247
- Space, A1, 17
- Square-free integers, 4.15
- Stable law, 377
- Standard normal distribution, 258
- State space, 111
- Stationary distribution, 124
- Stationary increments, 499
- Stationary probabilities, 124
- Stationary process, 363, 494
- Stationary sequence of random variables, 363
- Stationary transition probabilities, 111
- Stieltjes integral, 228
- Stirling formula, 27.18
- Stochastic arithmetic, 2.18, 4.15, 4.16, 5.19, 6.16, 18.17, 25.15, 393, 30.9, 30.10, 30.11
- Stochastic matrix, 112
- Stochastic process, 298, 308, 482
- Stopping time, 99, 133, 464, 465, 508
- Strong law of large numbers, 8, 9, 11, 85, 6.8, 282, 312, 27.20
- Strong Markov property, 508
- Subadditivity:
 - countable, 25, 162
 - finite, 24, 162
- Subfair game, 92, 102
- Submartingale, 462
- Subset, A1
- Subsolution, 8.5
- Sufficient subfield, 450
- Superharmonic function, 134
- Supermartingale, 462
- Support line, A33
- Support of measure, 23, 161
- Symmetric difference, A1
- Symmetric random walk, 113, 35.10
- Symmetric stable law, 378
- System, 111
- Tail σ -field, 63, 287, 496
- Tarski theorem, 3.8
- Taylor series, A29, 292
- Thin cylinders, 27
- Three series theorem, 290
- Tightness, 336, 380
- Timid play, 108
- Tonelli theorem, 234
- Total variation, 421
- Trajectory of process, 493
- Transformation of measures, 185, 215
- Transient, 117, 120
- Transition probabilities, 111
- Translation invariance, 45, 172
- Triangular array, 359
- Triangular distribution, 348
- Trifling set, 1.3, 1.9, 3.15
- Type, 193
- Unbiased estimate, 251, 454
- Uncorrelated random variables, 7, 277
- Uniform distribution, 258, 348
- Uniform distribution modulo 1, 328, 352
- Uniform integrability, 216, 338
- Uniformly equicontinuous, 26.15
- Uniqueness of extension, 36, 42, 163
- Uniqueness theorem for characteristic functions, 346, 382
- Uniqueness theorem for moment generating functions, 147, 284, 26.7, 390
- Unit interval, 51
- Unit mass, 24
- Upcrossing, 467
- Upper integral, 204, 228
- Upper semicontinuous, 29.1
- Upper variation, 421
- Utility function, 7.12
- Vague convergence, 371
- Value of payoff function, 134
- van der Waerden function, 31.18

- Variance, 78, 275
- Version of conditional expected value, 445
- Version of conditional probability, 430
- Vieta formula, 1.7

- Wald equation, 22.8
- Weak convergence, 190, 327, 378
- Weak law of large numbers, 5, 11, 86, 284
- Weierstrass approximation theorem, 87, 26.19

- Weierstrass M-test, 210, A28
- Weiner process, 498
- With probability 1, 60

- Zeroes of Brownian motion, 507
- Zero-one law, 63, 117, 8.35, 286, 22.12, 314, 496, 37.4
- Zorn's lemma, A8